

<http://www.phei.com.cn>

基础医学统计设计与数据分析

© 胡良平 余红梅 高 辉 主编



统计 分析 教材



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY



胡良平 教授，博士生导师，
现任军事医学科学院研究生
部医学统计学教研室主任，
生物医学统计学咨询中心主
任，国际一般系统论研究会
中国分会概率统计系统专业
理事会常务理事，第九届中
国现场统计研究会理事，中
国生物医学统计学会副会长，
《中华医学杂志》等10余种

杂志编委，北京大学口腔医学院客座教授，国家食品药
品监督管理局评审专家；主编统计学专著37部，参编统
计学专著10部；发表第一作者学术论文220余篇，发表
合作论文130余篇，获军队科技成果和省部级科技成果
多项；参加并完成三项国家标准的撰写工作；参加两项
国家科技重大专项课题研究工作；在从事统计学工作的
近30年中，为几千名研究生、医学科研人员、临床医生
和杂志编辑讲授生物医学统计学，在全国各地作统计学
学术报告100余场，举办几十期全国统计学培训班，培
养多名统计学专业硕士和博士研究生；近年来，参加国
家级新药和医疗器械项目评审数十项，参加50多项全军
“十二五”重大重点课题的统计学检查工作；归纳并提
炼出有利于透过现象看本质的“八性”和“八思维”的
统计学思想；独创了逆向统计学教学法和三型理论。擅
长科研课题的科研设计、复杂科研资料的统计分析与SAS
实现，以及各种层次的统计学教学培训和咨询工作。

基础医学统计设计与数据分析

◎ 胡良平 余红梅 高 辉 主编

◎ 刘红波 王 霞 副主编



统计分析教材

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书概括地总结了各类期刊论文中统计学应用的现状和在统计学应用方面存在的问题,先以 10 章的篇幅列举了各种期刊论文中时常出现的误用和滥用的典型案例,对案例中出现的差错逐一进行辨析与释疑,从反面引起广大读者的警觉,目的是用“反面刺激”有效提高广大读者识别统计学应用中可能出现的各种错误的能力,以期达到“吃一堑,长一智”之学习功效;接着以 5 章的篇幅从正面比较系统地介绍统计学的核心内容,内容涉及科研设计、统计表达与描述、统计分析方法合理选择和 SAS 实现等关键技术,以“提纲挈领、言简意赅”的写作手法,将统计学本身和使用方法的精华呈现在读者的面前。

本书适合从事生物医学、临床等工作的学生(专科生、本科生、进修生、硕士和博士研究生)、科技工作者和临床医护人员学习和使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

基础医学统计设计与数据分析/胡良平,余红梅,高辉主编. —北京:电子工业出版社,2014.6

统计分析教材

ISBN 978-7-121-23194-0

I. ①基… II. ①胡… ②余… ③高… III. ①医学统计-统计分析-高等学校-教材 IV. ①R195.1

中国版本图书馆 CIP 数据核字(2014)第 095777 号

策划编辑:秦淑灵

责任编辑:郝黎明

印 刷:

装 订:

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:787×1092 1/16 印张:19.75 字数:505.6 千字

版 次:2014 年 6 月第 1 版

印 次:2014 年 6 月第 1 次印刷

印 数:3000 册 定价:45.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zltz@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

编 委 会

主 编 胡良平 余红梅 高 辉

副 主 编 刘红波 王 霞

学术秘书 罗艳虹

编 委 （以单位和姓氏笔画为序）

山西医科大学 余红梅 罗艳虹 赵春妮

中国医科大学 刘红波

兰州大学 任晓卫

军事医学科学院 胡良平

徐州医学院 曾 平

第四军医大学 万 毅 王 霞

解放军 95969 部队卫生队 高 辉

前 言

从事医学基础研究的人们都不可避免地要运用统计学知识,因为他们要解决的问题、要攻克的课题都涉及多因素、多指标的生物医学和临床问题。然而,由于时间、精力、兴趣、数理基础、统计教材和教学方法等诸多方面的原因,他们在读书期间尽管学了不少一遍统计学(普通的科研人员平均学了约三遍统计学),但对于统计学的精髓和正确运用统计学的方法了解得还不够。因此,在从事各项工作(特别是科研工作)中,误用和滥用统计学的频率是相当高的。作为统计学工作者,我们深感不安。

为了有效地帮助从事医学基础研究和一切需要统计设计与数据分析的人们正确领悟统计思想,把握统计学各部分的要领,将科研、教学、咨询等一切工作做得更加科学、严谨、经济、可靠,笔者邀请山西医科大学公共卫生学院医学统计学教研室余红梅教授一道,组织山西医科大学、中国医科大学、兰州大学、徐州医学院、第四军医大学的多名统计学方面的年轻专家教授,针对数十种基础医学和医学综合类期刊中出现的问题,概括地总结了各类期刊论文中统计学应用的现况,运用逆向思维,先以10章的篇幅列举了各种期刊论文中时常出现的误用和滥用统计学的典型案例,对案例中出现的差错逐一进行辨析与释疑,从反面引起广大读者的警觉,从而有效提高广大读者识别统计学应用中可能出现的各种错误的能力;接着又以5章的篇幅从正面比较系统地介绍统计学的核心内容,即科研设计、统计表达与描述、统计分析方法的合理选择和SAS实现等关键技术,以“提纲挈领、言简意赅”的写作手法,将统计学本身和使用方法的精华呈现在读者的面前。

本书由军事医学科学院胡良平教授组织编写,其中第1章由余红梅编写,第2章由王霞编写,第3章由万毅编写,第4、8章由任晓卫编写,第5、10章由曾平编写,第6章由赵春妮编写,第7章由刘红波编写,第9章由罗艳虹编写,第11、12、13、14、15章由高辉、胡良平编写,全书由胡良平统稿、审校。

值得一提的是笔者的一位博士研究生高辉,他于2012年6月毕业后回到原先的工作单位,战备和训练工作极其繁忙,还挤出时间协助笔者审校了全书初稿,并按笔者的要求从正面撰写了本书的最后五章,为广大读者易于抓住统计学的精髓做出了很大的贡献。

由于笔者水平有限,书中难免会出现这样或那样的不妥,甚至错误之处,恳请广大读者不吝赐教,以便再版时修正。作者 E-mail: LPHU812@sina.com。

主编 胡良平
于北京军事医学科学院研究生部
生物医学统计学咨询中心

目 录

第 1 章 医院管理统计设计与数据分析	(1)
1.1 杂志近 5 年论文分析	(1)
1.2 统计设计中的常见问题及案例辨析	(2)
1.2.1 对照原则	(2)
1.2.2 随机化原则	(3)
1.2.3 重复原则	(5)
1.2.4 均衡原则	(5)
1.2.5 试验设计类型应用和判断中的错误	(6)
1.3 定量资料统计分析中的常见问题及案例辨析	(8)
1.3.1 未标明统计学方法	(8)
1.3.2 两独立样本的 t 检验样本例数未知可能导致结果偏差	(8)
1.3.3 未考察参数检验条件, 直接选用非参数检验	(8)
1.3.4 统计推断结果解释有误	(9)
1.3.5 误用成组设计定量资料 t 检验处理配对设计定量资料	(9)
1.3.6 方差分析结果的解释有误	(10)
1.3.7 误用 t 检验处理单因素多水平设计定量资料	(11)
1.3.8 资料对比未考虑重要非试验因素的影响	(12)
1.4 定性资料统计分析中的常见问题及案例辨析	(12)
1.4.1 未进行统计分析直接得出结论	(12)
1.4.2 未标明统计学方法	(13)
1.4.3 误用 χ^2 检验分析有序列联表资料	(14)
1.4.4 误用 t 检验处理定性资料	(15)
1.4.5 误用 χ^2 检验回答一致性问题	(15)
1.4.6 误用一般 χ^2 检验说明趋势问题	(16)
1.5 相关与回归分析中的常见问题及案例辨析	(17)
1.5.1 不同相关分析方法应用条件上混淆不清	(17)
1.5.2 过小的 r 值在下肯定相关结论时所遇到的问题	(18)
1.5.3 用单因素分析取代多因素分析	(19)
1.6 高级统计分析方法中的常见问题及案例辨析	(19)
1.6.1 主成分分析——指标选择条件不满足	(19)
1.6.2 重测信度——评价指标有误	(20)
1.6.3 因子分析——公共因子个数选择	(20)
1.6.4 结构方程模型——相关系数大于 1	(20)
1.7 统计结果报告中的常见问题及案例辨析	(21)
1.7.1 统计表使用的错误辨析与释疑	(21)
1.7.2 统计图使用的错误辨析与释疑	(23)

1.8 本章内容小结	(25)
参考文献	(25)
第2章 危重病急救医学统计设计与数据分析	(26)
2.1 急救医学研究概述	(26)
2.2 临床试验研究中常见的统计学问题及案例辨析	(27)
2.3 观察性研究中常见的统计学问题及案例辨析	(32)
2.4 试验研究中常见的统计学问题及案例辨析	(35)
2.5 本章常见的统计学问题小结	(37)
2.5.1 研究设计中常见的统计学问题	(38)
2.5.2 统计描述和统计推断中常见的统计学问题	(38)
2.5.3 统计结果解释中常见的统计学问题	(39)
参考文献	(39)
第3章 医学综合	(40)
3.1 《中华医学杂志》论文分析	(40)
3.2 统计设计中的常见问题及案例辨析	(41)
3.2.1 研究设计的类型、要素与原则	(41)
3.2.2 研究对象的选择与代表性	(44)
3.2.3 研究设计类型应用	(45)
3.3 统计分析中的常见问题及案例辨析	(46)
3.3.1 资料统计分析方法的合理选用	(46)
3.3.2 误用 t 检验分析多因素设计资料	(47)
3.3.3 误用单因素方差分析处理两因素资料	(48)
3.3.4 误用单因素方差分析处理三因素资料	(49)
3.3.5 非参数检验与 Fisher 精确检验	(50)
3.3.6 相关与回归分析误用	(52)
3.4 统计结果报告中的常见问题及案例辨析	(53)
3.4.1 假设检验的结果表达	(53)
3.4.2 统计表使用的错误辨析与释疑	(54)
3.4.3 统计图使用的错误辨析与释疑	(54)
3.5 本章内容小结	(55)
参考文献	(56)
第4章 医学遗传学统计设计与数据分析	(57)
4.1 科研设计错误辨析	(57)
4.1.1 研究对象选择不当	(57)
4.1.2 样本含量太少	(58)
4.1.3 均衡原则方面存在的问题	(58)
4.2 定量资料统计分析中存在的问题	(59)
4.2.1 误用 t 检验处理单因素多水平设计定量资料	(59)
4.2.2 误用 t 检验和单因素多水平设计定量资料方差分析处理多因素析因设计定量资料	(61)
4.2.3 误用 t 检验和单因素方差分析处理多因素重复测量设计定量资料	(62)
4.2.4 本节相关知识简要回顾与小结	(64)

4.3	定性资料统计分析中存在的问题	(64)
4.3.1	多重比较的校正问题	(64)
4.3.2	分层分析缺少 CMH χ^2 检验	(69)
4.3.3	误用 χ^2 检验处理结果变量为有序变量的单向有序列联表资料	(70)
4.3.4	本节相关知识简要回顾与小结	(71)
4.4	多变量回归分析方面存在的问题	(71)
4.4.1	多重线性回归方面存在的问题	(71)
4.4.2	Logistic 回归分析方面存在的问题	(72)
4.4.3	本节相关知识简要回顾与小结	(75)
	参考文献	(75)
第5章	疫苗和免疫统计设计与数据分析	(76)
5.1	计数资料模型	(76)
5.1.1	计数资料的特点	(76)
5.1.2	Poisson 分布与 Poisson 回归	(76)
5.1.3	负二项分布与负二项回归	(78)
5.1.4	过度离散检验	(78)
5.2	Meta 分析	(80)
5.2.1	Meta 分析的概念	(80)
5.2.2	Meta 分析的基本步骤	(80)
5.2.3	固定效应和随机效应模型	(81)
5.2.4	贝叶斯 Meta 分析	(81)
5.3	实例分析	(84)
5.3.1	计数资料分析	(84)
5.3.2	贝叶斯 Meta 分析	(86)
5.3.3	SAS 程序	(89)
	参考文献	(92)
第6章	高血压病统计设计与数据分析	(94)
6.1	论著统计方法应用情况	(94)
6.2	试验设计中的常见问题及案例辨析	(94)
6.3	统计分析中的常见问题及案例辨析	(98)
6.3.1	偏态分布误按正态分布处理	(98)
6.3.2	未标明所采用的统计分析方法	(98)
6.3.3	误用统计描述结果作出统计推断解释	(99)
6.3.4	误用单因素方差分析处理重复测量资料	(99)
6.3.5	误用随机区组设计方差分析处理完全随机设计资料, 并进行两两比较	(99)
6.3.6	误用 t 检验和单因素多水平设计定量资料方差分析处理具有一个重复测量两因素 设计定量资料	(100)
	参考文献	(101)
第7章	病理研究中的统计设计与数据分析	(102)
7.1	杂志近5年论文分析	(102)
7.2	试验设计和资料收集存在的问题	(103)

7.2.1	设计不全面, 资料收集混乱, 分析不清楚	(103)
7.2.2	受试对象少, 样本代表性差	(103)
7.3	定量资料统计分析中存在的问题	(104)
7.4	定性资料统计分析中存在的问题	(105)
7.4.1	误用一般四格表资料 χ^2 检验分析配对设计四格表资料	(105)
7.4.2	误用多次一般四格表资料 χ^2 检验分析多组比较定性资料	(106)
7.5	等级资料统计分析中存在的问题	(107)
7.5.1	误用 χ^2 检验分析等级资料	(107)
7.5.2	误用多次两样本秩和检验分析单因素多水平设计有序资料	(108)
7.5.3	误用 Mann-Whitney U 检验分析配对设计资料	(108)
7.5.4	误用 Wilcoxon 秩和检验分析配对设计有序资料	(109)
7.6	预后分析常见错误	(109)
7.6.1	分析结果意义不明确	(109)
7.6.2	结果描述过于简单, 表述不清	(110)
7.7	资料表达与结果描述中存在的统计学错误	(110)
7.7.1	错误定义检验水准	(110)
7.7.2	统计量标识不准确	(111)
7.7.3	P 值表达错误	(111)
7.7.4	相关性分析符号使用混乱	(111)
7.7.5	图表错误	(112)
7.8	统计学分析方法小结和 SPSS 菜单驱动的实现	(113)
7.8.1	定量资料的分析方法	(113)
7.8.2	定性资料的分析方法	(113)
7.8.3	有序资料的分析方法	(114)
7.8.4	多因素分析	(114)
7.8.5	用 SPSS 软件(菜单驱动法)实现常用统计分析	(115)
	参考文献	(116)
第 8 章	细胞与分子免疫学统计设计与数据分析	(117)
8.1	统计设计错误辨析	(117)
8.1.1	对照不全和采用的统计学分析方法错误	(117)
8.1.2	随机原则方面存在的问题	(119)
8.2	定量资料统计分析中存在的问题	(119)
8.2.1	误用成组设计定量资料 t 检验处理配对设计定量资料	(119)
8.2.2	误用 t 检验处理单因素多水平设计定量资料	(120)
8.2.3	误用单因素多水平设计定量资料方差分析处理多因素析因设计定量资料	(121)
8.2.4	误用 t 检验处理具有一个重复测量单因素设计定量资料	(123)
8.2.5	误用 t 检验处理具有一个重复测量两因素设计定量资料	(124)
8.2.6	误用 t 检验和单因素多水平设计定量资料方差分析处理多因素重复测量设计 定量资料	(125)
8.2.7	本节相关知识简要回顾与小结	(128)
8.3	定性资料统计分析中存在的问题	(129)
8.3.1	多重比较的校正问题	(129)

8.3.2	误用一般 χ^2 检验实现趋势分析	(131)
8.3.3	误用 χ^2 检验处理结果变量为有序变量的单向有序列联表资料	(132)
8.3.4	误用 χ^2 检验处理双向有序列联表资料	(133)
8.3.5	本节相关知识简要回顾与小结	(134)
8.4	相关与回归分析方面存在的问题	(134)
8.4.1	误用均值间差异比较处理相关性问题	(134)
8.4.2	Cox 回归分析方面存在的问题	(135)
8.4.3	本节相关知识简要回顾与小结	(136)
	参考文献	(136)
第9章	血液学统计设计和数据分析	(137)
9.1	杂志近5年论文分析	(137)
9.2	科研设计中的常见问题及案例辨析	(138)
9.2.1	回顾性描述资料及未进行统计分析直接下结论	(138)
9.2.2	组间均衡性差	(140)
9.2.3	样本例数太少	(140)
9.2.4	对照不全	(141)
9.2.5	本节相关知识简要回顾	(141)
9.3	定量资料统计分析中的常见问题及案例辨析	(142)
9.3.1	忽略参数检验的前提条件	(142)
9.3.2	误用 t 检验处理单因素多水平设计定量资料	(143)
9.3.3	误用析因设计定量资料方差分析和成组设计定量资料 t 检验处理具有一个重复测量两因素设计定量资料	(144)
9.3.4	未用统计方法分析具有一个重复测量三因素设计定量资料	(145)
9.3.5	未用统计方法分析两因素析因设计定量资料	(146)
9.3.6	误用单因素设计定量资料方差分析处理两因素析因设计定量资料	(147)
9.3.7	本节相关知识简要回顾	(147)
9.4	定性资料统计分析中的常见问题及案例辨析	(148)
9.4.1	计算率时分母过小不合适	(148)
9.4.2	误用 Fisher 确切概率法处理单向有序的列联表资料	(149)
9.4.3	误用 χ^2 检验分析结果变量为有序变量的列联表资料	(149)
9.4.4	误用 χ^2 检验取代多重 logistic 回归分析	(149)
9.4.5	本节相关知识简要回顾	(150)
9.5	相关回归分析中的常见问题及案例辨析	(151)
9.5.1	性质不同的数据合并进行相关分析	(151)
9.5.2	相关分析未画散点图及相关系数过小	(151)
9.5.3	用单因素分析取代多因素分析	(152)
9.5.4	筛选变量的方法有误	(153)
9.5.5	本节相关知识简要回顾	(154)
9.6	统计结果报告中的常见问题及案例辨析	(154)
9.6.1	统计结论写法有误	(154)
9.6.2	统计表有误, P 值一般写在表外	(155)
9.6.3	用“均值 \pm 标准误差”描述正态分布资料的分布特征	(156)

9.6.4	用“均值 \pm 标准差”描述偏态分布资料的分布特征	(156)
9.6.5	误用条图表达适合用线图表达的资料——横轴上的变量为时间且横轴 刻度表达有误	(156)
9.6.6	误用条图表达适合用线图表达的资料——横轴上的变量为浓度	(157)
9.6.7	本节相关知识简要回顾	(157)
9.7	本章内容小结	(158)
	参考文献	(158)
第 10 章	连续型随机变量概率分布与基因数据分析	(159)
10.1	常用统计分布	(159)
10.1.1	正态分布	(159)
10.1.2	χ^2 分布	(160)
10.1.3	t 分布	(160)
10.1.4	F 分布	(162)
10.2	置信区间和 P 值	(162)
10.2.1	基本概念	(162)
10.2.2	区间估计	(162)
10.2.3	P 值	(163)
10.2.4	贝叶斯统计和频率统计	(164)
10.2.5	置信区间和 P 值的频率统计解释	(164)
10.3	微阵列差别表达分析和多重假设检验	(166)
10.3.1	P 值和假阳性率	(166)
10.3.2	多重假设检验和总体错误率	(167)
10.3.3	错误发现率	(168)
10.3.4	贝叶斯错误发现率	(168)
10.3.5	FDR 和总体错误率的关系	(169)
10.3.6	对相关的敏感程度	(169)
10.3.7	无效分布	(169)
10.3.8	软件执行	(171)
10.4	实例分析和结果报告	(171)
10.4.1	前列腺癌微阵列数据实例	(172)
10.4.2	t 检验和总体错误率	(172)
10.4.3	错误发现率控制	(172)
10.4.4	错误发现率估计	(173)
10.4.5	结果报告	(173)
10.4.6	前列腺癌 FDR 分析的程序	(173)
	参考文献	(175)
第 11 章	科研设计方法及要点的把握	(178)
11.1	科研设计概述	(178)
11.1.1	科研设计的概念	(178)
11.1.2	科研设计整体架构	(178)
11.2	科研设计要点	(179)
11.2.1	基本常识与专业知识在科研设计中的作用	(179)

11.2.2	统计研究设计及其内在关系	(179)
11.2.3	完善的科研设计的标志以及获得完善科研设计方案的途径	(180)
11.3	试验设计要点	(181)
11.3.1	广义、狭义及简易型试验设计分别是什么	(181)
11.3.2	试验设计三要素及把握的要领	(182)
11.3.3	试验设计四原则及把握的要领	(190)
11.3.4	试验设计类型及把握的要领	(195)
11.3.5	如何进行试验过程中的质量控制	(199)
11.4	临床试验设计要点	(200)
11.4.1	临床试验中的法律与法规简介	(200)
11.4.2	临床试验的特点与分期及伦理道德问题	(200)
11.4.3	如何制定出合理的诊断、纳入和排除标准	(202)
11.4.4	如何确定主要疗效与安全性指标	(203)
11.4.5	如何遵循随机、对照、重复和均衡原则	(203)
11.4.6	如何实施盲法与降低偏性及多中心临床试验的中心效应	(207)
11.4.7	如何正确把握临床试验中的四种比较类型	(208)
11.5	调查设计要点	(208)
11.5.1	如何正确把握调查设计中的几个关键环节	(208)
11.5.2	常用抽样方法	(210)
11.5.3	如何进行调查质量控制	(212)
11.5.4	如何编制一份具有可操作性的调查表	(213)
11.5.5	调查资料信度与效度的评价及 SAS 实现	(214)
	参考文献	(215)
第 12 章	制订试验设计方案的关键技术	(216)
12.1	随机化的种类和 SAS 实现方法	(216)
12.1.1	随机化的体现	(216)
12.1.2	常用随机抽样方法的种类与 SAS 实现	(216)
12.1.3	常用随机分组方法的种类与 SAS 实现	(218)
12.2	样本含量和检验效能估计的 SAS 实现	(223)
12.2.1	估计样本含量的意义	(223)
12.2.2	确定样本含量时应具备的条件	(224)
12.2.3	估计总体均值时样本含量估计的 SAS 实现	(224)
12.2.4	估计总体率时样本含量估计的 SAS 实现	(225)
12.2.5	单组设计均值与率的检验时样本含量估计的 SAS 实现	(226)
12.2.6	配对设计均值与率的检验时样本含量估计的 SAS 实现	(227)
12.2.7	成组设计均值与率的差异性检验时样本含量估计的 SAS 实现	(229)
12.2.8	单因素多水平设计均值与率的检验时样本含量估计的 SAS 实现	(230)
12.2.9	检验效能分析概述	(231)
12.2.10	单组、配对设计定量资料假设检验时检验效能的 SAS 实现	(231)
12.2.11	成组设计均值与率的差异性检验时检验效能的 SAS 实现	(232)
12.3	多因素试验设计类型的构造及辨析	(233)
12.3.1	随机区组设计和具有一个重复测量的单因素设计	(233)

12.3.2	平衡不完全随机区组设计与双因素无重复试验设计	(234)
12.3.3	拉丁方设计与交叉设计	(235)
12.3.4	嵌套(系统分组)设计、裂区(分割)设计及重复测量设计	(236)
12.3.5	析因设计及含区组因素的析因设计	(239)
12.3.6	多因素试验设计类型的辨析	(241)
12.4	如何用 SAS 产生多因素试验设计类型	(245)
12.4.1	试验设计方案包含哪些内容	(245)
12.4.2	用 SAS 实现随机区组设计	(246)
12.4.3	用 SAS 实现拉丁方设计	(247)
12.4.4	用 SAS 实现 2×2 交叉设计	(248)
12.4.5	用 SAS 实现 3×3 交叉设计	(250)
12.4.6	用 SAS 实现析因设计	(251)
12.4.7	用 SAS 实现含区组因素的析因设计	(252)
12.4.8	用 SAS 实现平衡不完全区组设计	(254)
	参考文献	(255)
第 13 章	统计表达与描述的关键技术	(256)
13.1	以简化形式表达多组定量资料的关键技术	(256)
13.1.1	用“ $\bar{x} \pm s$ ”的形式表达析因设计定量资料	(256)
13.1.2	用“ $M(Q_1 \sim Q_3)$ ”的形式表达裂区设计定量资料	(257)
13.2	编制常用统计表的关键技术	(257)
13.2.1	表达定性资料的统计表	(257)
13.2.2	表达原始定量资料的统计表	(262)
13.3	绘制用于直观判断定量变量分布情况的统计图的关键技术	(263)
13.3.1	直方图(Histogram)应用场合及实例	(263)
13.3.2	P-P 图、Q-Q 图及概率图应用场合及实例	(264)
	参考文献	(266)
第 14 章	统计分析方法合理选择的关键技术	(267)
14.1	合理选择统计分析方法的一般步骤	(267)
14.2	合理选择统计分析方法处理一元定量资料的关键技术	(268)
14.3	合理选择统计分析方法处理一元定性资料的关键技术	(273)
14.4	合理选择统计分析方法处理多元定量资料的关键技术	(279)
14.5	合理选择统计分析方法处理多元定性资料的关键技术	(281)
	参考文献	(282)
第 15 章	如何用 SAS 实现常用统计分析的关键技术	(283)
15.1	用 SAS 实现统计分析的简便快捷的方法	(283)
15.2	用 SAS 实现一元定量资料的统计分析	(284)
15.3	用 SAS 实现一元定性资料的统计分析	(289)
15.4	用 SAS 实现多元定量资料的统计分析	(294)
15.5	用 SAS 实现多元定性资料的统计分析	(296)
	参考文献	(296)
附录 A	胡良平统计学专著及配套软件简介	(297)

第1章 医院管理统计设计与数据分析

《中华医院管理杂志》主要收录了有关医院管理多个方面的内容，如科研管理、护理管理、药事管理、精神卫生管理、人力资源管理、医院绩效管理、经济管理、设备管理、信息管理、服务管理、医疗纠纷管理、工作研究、经验交流、卫生保健、健康管理、突发事件管理、急诊救助管理、医院文化、医德医风、思想政治工作、医患关系等。

本章内容包括该杂志近5年论文(主要指论著)分析、统计设计、统计分析及统计结果报告中常见问题及案例辨析。

1.1 杂志近5年论文分析

《中华医院管理杂志》在2007—2011年间共发表1570篇论著，其中有统计分析内容的为513篇，约占1/3；513篇论著中，调查研究占绝大多数(92%)；采用某种统计推断方法的约占1/2。五年间该规律基本稳定，详见表1-1。

表 1-1 2007—2011 年《中华医院管理杂志》分类汇总 I——总况

分 类	分 组	篇数(%)					
		2007 年	2008 年	2009 年	2010 年	2011 年	合计
统计方法	无	256(74%)	143(61%)	226(70%)	228(65%)	204(66%)	1057(67%)
	有	89(26%)	93(39%)	99(30%)	125(35%)	107(34%)	513(33%)
	合计	345(100%)	236(100%)	325(100%)	353(100%)	311(100%)	1570(100%)
研究类型	调查研究	85(95%)	88(95%)	92(93%)	108(86%)	99(92%)	472(92%)
	试验研究	4(5%)	5(5%)	7(7%)	17(14%)	8(8%)	41(8%)
	合计	89(100%)	93(100%)	99(100%)	125(100%)	107(100%)	513(100%)
统计推断	无	53(60%)	42(45%)	51(51%)	60(48%)	56(52%)	262(51%)
	有	36(40%)	51(55%)	48(49%)	65(52%)	51(48%)	251(49%)
	合计	89(100%)	93(100%)	99(100%)	125(100%)	107(100%)	513(100%)

按照统计推断方法分类，513篇论著中，300篇次采用基本统计推断方法，其中 t 检验、方差分析、基于二项分布和Poisson分布的统计分析、 χ^2 检验、秩和检验、简单回归与相关分析的依次分别占17%、15%、2%、33%、9%和24%，详见表1-2。高级统计推断方法中，涉及因子分析、结构方程模型、综合评价、时间序列分析、聚类分析或其他方法，但是数量极少。

表 1-2 2007—2011 年《中华医院管理杂志》分类汇总 II——基本统计推断方法

方 法	篇数(%)					
	2007 年	2008 年	2009 年	2010 年	2011 年	合计
t 检验	6(17%)	12(23%)	10(21%)	12(19%)	10(15%)	50(17%)
方差分析	6(17%)	9(18%)	11(23%)	7(11%)	11(16%)	44(15%)
二项、Poisson 分布	0(0%)	0(0)	0(0)	1(2%)	4(6%)	5(2%)
χ^2 检验	21(58%)	21(41%)	21(44%)	20(32%)	17(25%)	100(33%)
秩和检验	5(14%)	3(6%)	6(12%)	8(13%)	6(7%)	28(9%)
简单回归与相关	13(36%)	16(31%)	9(19%)	14(23%)	21(30%)	73(24%)
合计	51(100%)	61(100%)	57(100%)	62(100%)	69(100%)	300(100%)

1.2 统计设计中的常见问题及案例辨析

试验设计中有以下三个主要知识点：即试验设计的三要素、试验设计的四原则和试验设计类型。所谓“三要素”，是指试验研究中不可缺少的三个基本元素，即“受试对象、影响因素（包括试验因素和非试验因素）和试验效应”。所谓“四原则”，就是指试验研究必须遵循的四个基本原则，即“对照、重复、随机、均衡”原则。所谓试验设计类型，就是安排因素及其水平的一种架构，以便实施具体试验并合理选择统计分析方法处理数据。具体地说，就是根据具体试验研究中将涉及的试验因素的个数和拟加以控制的重要非试验因素的个数及其水平数等多方面的情况综合考虑，选择合适的方案安排多个因素，以便能尽可能准确地估计各因素及其交互作用的大小。正确运用这“三个知识点”辨析医学科研中试验设计方面存在的问题，可提高医学科研工作的科学性和严谨性。

1.2.1 对照原则

【案例 1-1】 耐甲氧西林金黄色葡萄球菌(MRSA)是医院感染最主要的致病菌之一。自 1961 年首次报道第 1 株 MRSA 以来，其临床检出率在全球范围内呈显著上升趋势，尤其在重症监护病房。对象：2007 年 1 月至 2008 年 12 月，检出的所有住院患者 MRSA 菌株中，符合卫生部医院感染诊断标准的病例（剔除社区获得性感染、外院获得性感染病例）。资料收集包括患者年龄、基础疾病、感染日期、感染部位。2007 年 1 月至 12 月获得性 MRSA 病例为常规组，2008 年 1 月至 12 月为干预组。

【案例辨析】 (1) 该研究的“三要素”：受试对象，2007 年 1 月至 12 月获得性 MRSA 病例；处理因素，单因素两水平，医院获得性耐甲氧西林金黄色葡萄球菌感染干预与否，它是一个具有两水平的试验因素，未涉及重要非试验因素；试验效应，MRSA 检出率。(2) 原作者直接选择 2007 年 1 月至 12 月医院获得性 MRSA 病例为常规组，2008 年 1 月至 12 月为干预组，MRSA 所致感染呈散发或爆发流行，如果在不清楚当时的 MRSA 总体感染情况下就直接加干预进行比较，这样是不具有说服力的。(3) 该试验研究对试验组采取干预措施，常规组未采取，但是即使不采取干预措施，医院对之前的常规组仍会有基本的控制措施，原作者并没有指出常规组是否为完全空白对照，或是采取的基本控制措施与干预措施的不同之处。

【案例释疑】 (1) 原作者在研究时应提供常规组与干预组两个时间段内的 MRSA 流行状况，分析两组除了干预措施施加与不施加以外影响院内感染的因素是否均衡，增强其可比性。若两组患者病情轻重相差悬殊，就无法正确评价干预措施的真实效果。(2) 原作者应尽量详细介绍常规组是否为空白对照，或介绍其采取的基本措施与干预措施之间的差异。

【案例 1-2】【情况 1】 目的：从编码使用频度角度考察国际疾病分类第十次修订本(ICD-10)北京临床版的应用状况，以利于拓展应用和发展。方法：对 ICD-10 北京临床版应用前(2006 年)、推广中(2007 年)和应用中(2008 年)编码使用频度差异情况进行统计、计算相似度并做比较分析。

【情况 2】 该院从 2003 年开展 PICC 技术至今已有 11 年，对 PICC 技术的应用实行了系统化管理。原作者通过比较系统化管理前后 PICC 置管数量和质量来讨论系统化管理的实施对 PICC 的影响。

【情况 3】 介绍了安徽省城乡医疗救助政策的实施情况及产生的社会效果，分析其中存在的问题，提出了完善医疗救助政策的建议。安徽省于 2004 年 6 月在农村先试点医疗救助政策，从 2005 年 6 月开始在全省农村全面实施。原作者通过比较措施实施之前即 2004 年之前与实施之后即 2009 年的医疗救助情况来评价安徽省医疗救助政策实施效果。

【情况4】 收集2007年9月至11月和2008年1月至3月药师日常调剂过程中审核发现的不合理处方,观察用量限制模块运行前、后各3个月的指标变化(2007年12月为用量限制模块嵌入时间)。

【情况5】 某院对全院进行了系统调查之后,针对现存问题进行了系统化、专业化的培训和管理,于2009年8月成立IV-EAM。该研究通过比较实施前后的输液质量来讨论,IV-EAM在提高输液安全质量方面起到了积极的作用。

【案例辨析】 以上研究均采用自身对照的方式,比较同一对象采取措施前后的效果,从而来评价实施措施产生的影响。研究中原作者在选择研究对象时都是直接收集采取措施前后医院的病例资料来比较,采取措施后医院通常有详细的随访资料,而之前的并不一定有,随机选取两阶段的病例直接比较,他们之间的均衡性没有考虑。再者,以上研究都是以一家医院为研究总体,其研究结果的说服力也是有限的。

【案例释疑】 以上研究可以在自身对照设计的同时,增加另一个可以与其均衡可比的医院做平行对照,该医院不采用相应措施,而采用了措施的医院为试验组,分别将试验组与对照组两阶段的效果之差作比较,从而来评价该措施的影响。

1.2.2 随机化原则

【案例1-3】 选取2008年1月至2010年5月施行体外循环心内直视手术后拟入ICU的住院患者113例,年龄25~80岁,无意识障碍、语言障碍,均为首次入住ICU。其中男性72例,女性41例,平均年龄(44.2 ± 1.7)岁。临床诊断:先天性心脏病8例,心脏瓣膜病67例,冠状动脉粥样硬化性心脏病35例,主动脉夹层3例。按入院日期单双号随机分组,将113例患者分为术前访视组62例和对照组51例,均在气管插管全麻下进行心脏手术。两组性别、年龄、病种、术式等差异无统计学意义($P > 0.05$)。

【案例辨析】 该研究的“三要素”,受试对象:2008年1月至2010年5月施行体外循环心内直视手术后拟入ICU的住院患者113例,年龄25~80岁,无意识障碍、语言障碍,均为首次入住ICU;处理因素:术前访视实施与否;试验效应:患者护理效果、手术后焦虑水平、机械通气时间、入住ICU时间。

原作者的随机分组方式为按入院日期单双号随机分组,将113例患者分为术前访视组62例和对照组51例。患者来医院就诊,在患病严重程度、患病时间等重要非试验因素方面不一定是均衡的。如果按日期单双号分组,最后可能使两组在某些重要非试验因素方面相差悬殊,组间缺乏均衡性。

【案例释疑】 针对此类问题,有一种简便易行的随机化分组方法叫“按不平衡指数最小的分配原则进行随机化”。做法如下:根据专业知识选取几个重要的非试验因素,假定一个是患者的“性别”,一个是患者的“病情”。将先来的两位患者在试验组和对照组各放一个人,记下他们的性别和病情,记分的方法是每个因素的每个水平出现一次计一分,计算两组各因素对应水平的得分之差的绝对值,最后求出绝对值之和,称此“和”为患者两个重要的非试验因素上的不平衡指数,若再来第三个患者,分别将此患者放入试验组、对照组各一次,每次都根据该患者的性别和病情累加到原来患者的基础之上,可以得到两个不平衡指数,取不平衡指数最小的那种分组方法,这样第三位患者的分组就定下来了,用同样的方法去分配以后的患者,直到两组都累积到事先规定的样本含量时就停止。

【案例1-4】 对医学模拟培训中心提高住院医师临床技能效果进行定量评价,根据《广东省三级医院管理督查评价细则》相关规定设定培训的技能项目和评分标准,选择了徒手心肺复苏(CPR)、胸腹(膜腔)穿刺、中心静脉置管、气管插管、呼吸机使用五项基本技能作为培训和考核项

目。受训住院医师依工号随机分为培训组和对照组,根据住院医师所从事的专业来培训和考核不同的项目。每项技能组研究人数从10人至138人不等。

【案例辨析】 该研究的“三要素”,受试对象:“三甲”医院临床、检验、药剂、影像专业的住院医师,其学历、年龄、性别不限,技师和护师不在本研究范围;处理因素:住院医师规范化培训与否;试验效应:徒手心肺复苏(CPR)、胸腹(膜腔)穿刺、中心静脉置管、气管插管、呼吸机使用五项基本技能的成绩。

原作者在选择受试对象时其学历、年龄、性别不限,受训住院医师依工号随机分为培训组和对照组,原作者没有考虑受试对象的学历、年龄、性别等重要非试验因素就直接按工号随机分组,显然没有遵循试验设计的随机分组原则。

【案例释疑】 应该采用分层随机化进行分组,也可采用简便易行的随机化分组方法“按不平衡指数最小的分配原则”分组。

【案例1-5】 368例入选CABG患者被随机分为健康管理组193例和对照组175例,进行随访研究。对照组:术后按传统康复方式随诊,即带口服药出院,遵医嘱3~6个月复查。患者多根据自己的实际情况选择复查时间和医院,患者主动,医生被动。随诊流程再造组:实施“大医院医生——社区医疗单位——患者”点对点管理,医生主动,患者被动或主动,培训随诊方法和干预措施,实行网络管理,至少每1~3个月随诊1次。评价内容包括血压、血脂、血糖及健康状况。

【案例辨析】 该试验研究的“三要素”,受试对象:2004年8月至2005年8月北京同仁医院、北京健宫医院、上海同济医院等冠状动脉旁路移植术患者368例。纳入标准:诊断明确的CABG手术的患者,同意接受指定量表评估;排除标准:排除微创切口搭桥、搭桥合并主动脉瓣和/或二尖瓣置换等患者。处理因素:实施冠状动脉搭桥术后患者随诊管理。试验效应:生活质量及血压、血糖、血脂的异常率。

原作者虽然在选择试验对象后进行随机分组,但我们知道冠状动脉旁路移植术患者年龄在45~83岁,不同的年龄段其术后的恢复状况会不同;再者评价指标采用血压、血脂、血糖等,这些指标都与患者的自身状况密切相关,所以分组时应该考虑他们本身是否患有相关疾病。

【案例释疑】 应该慎重考虑这些重要的非试验因素,然后再进行随机化分组。可考虑按重要非试验因素对受试对象进行分层随机化或在统计分析阶段采用多变量回归分析校正这些非试验因素对因变量的影响。

【案例1-6】 为北京市某综合医院提供为期1年的对医务人员实施员工帮助计划干预(EAP)服务,采用SF-36健康调查量表和JSS量表,对医务人员的生命质量和工作满意度进行调查。经过EAP服务培训后进行EAP实施效果评价,如医务人员的心理现状、实施EAP前后医务人员身心健康状况的改善及医务人员工作满意度的改善状况。

【案例辨析】 该研究的“三要素”:受试对象,北京市某综合医院医务人员400人;处理因素,对医务人员实施员工帮助计划干预与否;试验效应,医务人员的心理现状、实施EAP前后医务人员身心健康状况的改善及医务人员工作满意度的改善状况。

原作者在2007年5月随机抽取北京市某综合医院医务人员400人,随机分配到接受EAP服务组(干预组)和对照组,其中EAP服务组200人,对照组200人。虽然表面上进行随机分组,但很模糊,并没有考虑到这400人在性别、年龄、婚姻状况、学历、身体状况、工作环境等重要非试验因素上是否均衡可比,这些因素会对试验效应产生很大影响。原作者没有在干预措施实施以前考虑干预组与对照组是否具有可比性。

【案例释疑】 应该慎重考虑这些重要的非试验因素,然后再进行随机化分组。可考虑按重要

非试验因素对受试对象进行分层随机化或在统计分析阶段采用多变量回归分析校正这些非试验因素对因变量的影响。

1.2.3 重复原则

【案例1-7】 根据《广东省三级医院管理督查评价细则》相关规定设定培训的技能项目和评分标准,选择了徒手心肺复苏(CPR)、胸腹(膜腔)穿刺、中心静脉置管、气管插管、呼吸机使用五项基本技能作为培训和考核项目。受训住院医师依工号随机分为培训组和对照组,根据住院医师所从事的专业来培训和考核不同的项目。每项技能组研究人数从10人至138人不等。

表 1-3 培训组与对照组“五项技能”成绩比较讨论

项 目	分 组	人 数	平 均 分	标 准 差	t 值	P 值
CPR	培训组	39	88.86	4.46	1.61	>0.05
	对照组	138	82.58	9.34		
胸腹穿刺	培训组	21	96.67	3.15	2.77	<0.01
	对照组	12	88.08	6.37		
中心静脉置管	培训组	16	89.88	9.19	0.34	>0.05
	对照组	10	88.10	5.36		
气管插管	培训组	16	95.80	5.71	0.71	>0.05
	对照组	10	93.00	5.85		
呼吸机使用	培训组	16	90.40	9.30	2.26	<0.05
	对照组	10	82.38	5.39		

【案例辨析】 该研究的“三要素”,受试对象:“三甲”医院临床、检验、药剂、影像专业的住院医师,其学历、年龄、性别不限,技师和护师不在本研究范围;处理因素:对培训组住院医师进行规范化培训;试验效应:徒手心肺复苏(CPR)、胸腹(膜腔)穿刺、中心静脉置管、气管插管、呼吸机使用五项基本技能的成绩。

原作者在选择受试对象时其学历、年龄、性别不限,每项技能组研究人数从10人至138人不等。从表1-3可看出培训组与对照组“五项技能”成绩比较时各组的样本含量多在10~20之间,样本含量较少,其结果说服力不强。CPR指标两组比较时培训组为39例,对照组为138例,两组例数相差较大,这对其结果有较大影响。

【案例释疑】 应该按照统计设计进行样本含量估计,并在两组间进行均衡性检验,减小两组间人为造成的误差,提高检验效率。

1.2.4 均衡原则

【案例1-8】 试验者选取2009年1月至2010年4月,在该医院科室运用PICC置管术进行化疗的患者100例,对照组选取2008年1月至2009年1月实行电子档案跟踪管理之前的PICC置管患者100例。选取对象均为手术后继续在本病区进行化疗,化疗总疗程大于3个月的患者。其中女性48例,男性52例;年龄范围为21~85岁。病种包括直肠癌、结肠癌、胃癌、胰腺癌和乳腺癌。

【案例辨析】 该试验研究的“三要素”,受试对象:试验组选取2009年1月至2010年4月,在该医院科室运用PICC置管术进行化疗的患者100例,而对照组选取2008年1月至2009年1月实行电子档案跟踪管理之前的PICC置管患者100例;处理因素:实行电子档案跟踪管理与否;试验效应:相应的并发症,如意外导管、导管外滑、导管断裂、堵管、未定期维护。

原作者在选择受试对象时具有很大的随意性,仅以是否实行电子档案跟踪管理而各选出100例患者作为试验组和对照组。分组表面上是随机的,选取对象均为手术后继续在本病区接受化疗,

化疗总疗程大于3个月的患者。但各组患者可能在病情等某些重要的非试验因素上相差很大,如性别、年龄、病种、病情严重程度,以及生活方式等方面是否都具有可比性。该研究并没有通过“随机化”方法尽量降低这些重要非试验因素对观测结果的影响,同时也严重地违背了试验设计的“均衡原则”,其试验结果的可靠性必然受到影响。另外“其中女性48例,男性52例”,是否试验组与对照组女性与男性的人数之比例均是如此,此处有歧义。

【案例释疑】 针对此项临床研究,可考虑采用区组设计,一个试验因素为是否采用电子跟踪,另外将重要的非试验因素如年龄、病种、病情等进行配对,这样尽可能消除非试验因素的影响,精确地控制误差,分析出试验因素的效应,得到较为可靠的结果。

【案例1-9】 资料来源:从2008年8月开始在医院3个病区的某一病区实施脑血管病介入治疗临床路径。现以2008年8月至2009年3月该病区收治的患者为临床路径组,选取同时期另一病区收治的脑血管病介入治疗的患者为对照组。进入临床路径的患者50例,其中6例中途退出,实际应用44例,对照组77例。对脑血管病介入治疗临床路径实施前后的环节质量和终末质量,主要是住院天数、住院总费用、单项住院费用和患者满意度等进行对比分析。

【案例辨析】 该研究的“三要素”,受试对象:以2008年8月至2009年3月收治的脑血管病介入治疗的患者;处理因素:实施脑血管病介入治疗临床路径与否;试验效应:住院天数、住院总费用、单项住院费用和患者满意度。

原作者在选择受试对象时以2008年8月至2009年3月该病区收治的患者为临床路径组,选取同时期另一病区收治的脑血管病介入治疗的患者为对照组,所有日期均以入院日期进行分类。我们知道,在研究时选择医院患者为受试对象往往具有选择偏倚,而直接以不同病区收治的患者来划分试验组(路径组)和对照组显然是不合理的。

【案例释疑】 在分组时应该考虑两病区内患者的年龄、病情、经济状况、病区内的医护人员配置情况等重要非试验因素的均衡可比性,可分别在两个病区各随机选取病例再进行随机化分配路径组和对照组。

1.2.5 试验设计类型应用和判断中的错误

【案例1-10】 选择两组镊子罐,即无盖无菌干燥镊子罐及带盖无菌干燥镊子罐。按春、夏、冬三季分别采样,带盖镊子罐为试验组,干镊子罐为对照组,每组镊子罐5个,内放无菌持物镊。设计24小时内更换镊子罐的时间为4、6、8、12、24小时(各时间点上使用不同的罐子)。用无菌镊夹取无菌物品,夹取物品次数分别为4小时内15次、6小时内20次、8小时内35次、12小时内40次、24小时内45次后,按设计时间分别对无菌镊子罐前端、无菌镊子罐罐口及罐底进行采样,进行微生物学检测。在任何一个镊子罐的3个采样部位中,如果有1个采样部位培养结果为有菌生长,为阳性。治疗室每日通风紫外线照射消毒,随机监测治疗室空气洁净度及治疗室的温湿度。统计学处理:计数资料用例数,百分数描述。统计学分析各时点两样本率的比较用 χ^2 检验;条件不符合时用确切概率法,详见表1-4。所有统计在SAS 8.0软件中完成,结果如下。

【案例辨析与释疑】 该试验研究的“三要素”,受试对象:无菌干燥镊子罐;处理因素:是否带盖,测量部位,测量时间;试验效应:微生物学检测细菌培养的阳性例数和阳性率。

该研究设计是具有一个重复测量三因素设计,其中一个试验因素是“是否带盖”,另一个是“时间”,重复测量因素是“测量部位”。若将原表改为表1-5的形式,则容易辨出它所属的设计类型。

表 1-4 干镊子罐与带盖镊子罐不同时点不同部位微生物比较结果

	时间 (h)	阳性例数	阳性率 (%)	阳性例数	阳性率 (%)	χ^2	P 值
		干镊子罐 (n=25)		带盖镊子罐 (n=25)			
罐口	4	1	4	0	0	—	1.0000
	6	3	12	0	0	—	0.2300
	8	8	32	2	8	4.50	0.0300
	12	11	44	4	12	6.35	0.0100
	24	17	68	9	36	5.13	0.0200
罐底	4	1	4	0	0	—	1.0000
	6	3	12	0	0	—	0.2300
	8	6	24	0	0	—	0.0200
	12	12	48	1	4	12.58	0.0004
	24	20	80	3	12	23.27	0.0000
罐缘	4	0	0	0	0	—	—
	6	0	0	0	0	—	—
	8	0	0	0	0	—	—
	12	1	4	0	0	—	1.0000
	24	2	8	1	4	—	1.0000
合计	4	1	4	0	0	—	1.0000
	6	4	16	0	0	—	0.1100
	8	10	40	2	8	7.02	0.0081
	12	14	56	3	12	10.78	0.0010
	24	21	84	10	40	10.27	0.0014

表 1-5 镊子罐与带盖镊子罐不同时点不同部位微生物比较结果

是否带盖	测量时间 (h)	阳性例数		
		灌口	罐底	灌缘
带盖镊子罐 (n=25)	4	0	0	0
	6	0	0	0
	8	2	0	0
	12	4	1	0
	24	9	3	1
干镊子罐 (n=25)	4	1	1	0
	6	3	3	0
	8	8	6	0
	12	11	12	1
	24	17	20	2

【案例 1-11】为北京市某综合医院提供为期 1 年的对医务人员实施员工帮助计划干预 (EAP) 服务, 采用 SF-36 健康调查量表和 JSS 量表, 对医务人员的生命质量和工作满意度进行调查。选取 2007 年 5 月, 随机抽取北京市某综合医院医务人员 400 人, 随机匹配到接受 EAP 服务组 (干预组) 和对照组, 其中 EAP 服务组 200 人, 对照组 200 人。经过 EAP 服务培训后对试验组和对照组进行 EAP 实施效果评价。

【案例辨析与释疑】该研究的“三要素”, 受试对象: 北京市某综合医院医务人员 400 人; 处理因素: 实施员工帮助计划干预与否; 试验效应: 医务人员的心理现状、实施 EAP 前后医务人员身心健康状况的改善及医务人员工作满意度的改善状况。

原作者在分析实施效果时分析了医务人员的心理现状,医务人员对 EAP 服务的需求情况,干预组实施 EAP 前后医务人员身心健康状况的改善及医务人员工作满意度的改善状况,对照组在 EAP 服务培训后医务人员身心健康状况的改善及医务人员工作满意度的改善状况。显然,如此分析已忽略两组之间的对比,应该同时考虑两组在 EAP 实施前后的效果,比较两组干预措施实施前后的平均差值之间的差别是否具有统计学意义。

1.3 定量资料统计分析中的常见问题及案例辨析

1.3.1 未标明统计学方法

【案例 1-12】 目的:探索提高医院感染管理水平的综合干预策略。干预前后各类人员成绩比较,统计学方法未写出,结果得出:终末调查各类人员医院感染相关知识成绩比基线调查均明显提高(表 1-6)。

表 1-6 不同人员培训前后成绩比较

人 员	人 数	分 数	人 数	分 数	P 值
	培 训 前		培 训 后		
领导	21	61.60 ± 14.07	21	71.43 ± 13.08	0.034
医务人员	442	65.53 ± 12.66	559	78.78 ± 10.49	0.000
医院感染控制专职人员	28	58.56 ± 17.07	28	85.48 ± 8.00	0.000

【案例 1-13】 2006 年,为了深入开展“医院管理年”和创建人民满意医院活动,加强医务人员基本技能训练,规范临床操作,提高医疗服务质量。在北京市卫生局所属医院中,某院率先开展了以“三基、三严”为重点的岗位练兵活动。在岗位练兵竞赛结果与分析中,统计方法未写出,得出结果:医技科室与临床科室得分比较,无显著性差异($P > 0.05$)。

【案例辨析与释疑】 文中未标明使用了哪种统计分析方法,而直接给出 P 值范围,做出结论,缺乏说服力,可信度很低。

1.3.2 两独立样本的 t 检验样本例数未知可能导致结果偏差

【案例 1-14】 为了解女性乳腺癌患者信息需求及影响因素研究,于 2009 年 7 月到 2010 年 12 月采用方便抽样的方法抽取 2 家长沙市三级甲等医院作为调查点,再随机抽取符合条件的患者参与调查,进行现场调查问卷,采用 SPSS 17.0、Amos 7.0 建立数据库并进行统计处理,女性乳腺癌患者的总需求得分为 65.61 ± 20.854 ,且处于手术治疗阶段、年轻、收入高的患者需求强烈。提示不同的社会属性决定信息需求不同。

【案例辨析与释疑】 本文在不同年龄需求比较时,按照 50 岁为临界值将患者分为两组,使用两样本 t 检验比较不同分组的需求差异,青年组的需求得分为 65.604 ± 20.929 ,老年组的需求得分为 65.652 ± 20.746 ,即认为两者无统计学差异的结果不可靠。文中提到被调查者的年龄为 21 ~ 72 岁,而 36 ~ 50 岁的患者有 228 名,占有调查对象总人数的 70.59%,这可能会导致老年组的患者人数过少,代表性不够,从而得到与实际不符的结果。

1.3.3 未考察参数检验条件,直接选用非参数检验

【案例 1-15】 我国西部农村贫困家庭健康风险描述与特征研究中,采取典型抽样的方式,选择国家级贫困地区重庆市黔江区和贵州省贵定县作为样本点。在 11 个乡共抽取 1109 户农村

家庭作为调查样本,以调查问卷进行现场调查,回顾自调查日期的前1年的健康需求与利用情况。将得到的数据采用 SPSS 12.0 专业统计软件处理。结果表明贫困家庭成员2周患病率和慢性病患病率均显著性地高于非贫困家庭成员,但患病者治疗的比例显著性地低于非贫困家庭成员。

【案例辨析与释疑】 本文在贫困家庭和非贫困家庭成员住院花费情况比较时没有先对资料进行参数检验条件考察,而直接采用秩和检验,可能导致“将符合参数检验的资料采用了非参数检验,降低检验效能”的结果。应该先对数据进行考察,如果符合参数检验条件,则首选参数检验。如果不符合参数检验条件,再尝试变量变换,或采用非参数检验。

1.3.4 统计推断结果解释有误

【案例 1-16】 在不同个人特征的患者医疗服务期望的差异性分析中,选取温州市4所医院,对其中的350名患者进行问卷调查,借鉴利克特五分量表评价方式对服务期望中的地位打分,运用 SPSS 统计分析软件录入数据进行单因素方差分析,得出不同年龄及不同文化程度的患者医疗服务期望具有差异。

【案例辨析与释疑】 本文对不同年龄患者医疗服务的差异进行分析时,认为通过表1-7可以看出:60岁以上的患者对信任感的要求最高,其次是51~60岁年龄段的患者,解释有误。从表1-7中可看出21~30岁年龄段与31~40岁年龄段的患者对信任感的要求均高于51~60岁年龄段的患者,而51~60岁年龄段的患者对信任感的要求低于60岁以上年龄段的患者。无法得出51~60岁年龄段的患者仅比60岁以上的患者对信任感的要求低,而比其他组高。

表 1-7 不同年龄患者医疗服务期望的方差分析

调查内容	方差值	P值	患者类型	两组均数差值	P值
信任感	2.587	0.028	21~30-51~60	1.2292	0.015
			31~40-51~60	1.4840	0.004
			51~60-60以上	-1.6545	0.002
有效性	3.309	0.007	0~20-60以上	-2.0000	0.041
			21~30-51~60	1.4966	0.007
			31~40-51~60	1.8062	0.001
			51~60-60以上	-1.9600	0.001

1.3.5 误用成组设计定量资料 t 检验处理配对设计定量资料

【案例 1-17】 对医务人员实施员工帮助计划干预效果评价,为北京市某综合医院提供1年的EAP服务,采用SF-36健康调查量表和JSS量表,对医务人员的生命质量和工作满意度进行调查,用SPSS专业统计软件分析,得出工作压力是医务人员最主要的压力源,占63.7%。接受EAP服务组的SF-36量表及JSS量表中各维度(工作性质和上司维度除外)的得分显著高于对照组,差异有统计学意义($P < 0.01$)。说明医务人员的工作压力较大,对EAP的需求意愿强烈。EAP服务是维护医务人员心理健康、提高生命质量和工作绩效的有效手段,可以推广使用。

【案例辨析与释疑】 本文对于干预组EAP实施前后的比较采用成组设计定量资料 t 检验处理(表1-8)有误。对于干预组EAP实施前后的比较,属于自身前后对照的配对设计,应该首先考察每对数据的差值是否满足正态性,如果满足可采用配对设计定量资料 t 检验,如不满足正态性可采用配对设计定量资料 Wilcoxon 符号秩检验。

表 1-8 EAP 实施前后干预组的工作满意度量表得分情况 ($\bar{x} \pm s$)

工作要素	实 施 前	实 施 后	<i>t</i> 值	<i>P</i> 值
工作性质	13.92 ± 4.22	14.84 ± 2.77	-2.67	0.080
上司	15.02 ± 4.30	14.37 ± 2.10	1.54	0.125
同事	14.94 ± 2.88	16.44 ± 2.35	-5.61	0.000
福利	10.33 ± 3.42	14.46 ± 2.07	-15.59	0.000
沟通	13.41 ± 3.25	14.49 ± 3.15	-3.11	0.000
违章制度	11.65 ± 3.00	13.88 ± 2.74	-7.06	0.000
认同感	12.00 ± 3.51	15.12 ± 2.63	-10.16	0.000
报酬	10.25 ± 3.54	14.51 ± 2.01	-16.11	0.000
晋升	11.69 ± 3.66	14.16 ± 2.84	-7.52	0.000

1.3.6 方差分析结果的解释有误

【案例 1-18】 公立医院医生心理资本特征研究中,按照分层随机抽样的原则,对潍坊市 11 所公立医院医生进行调查,其中一级医院 6 所、二级医院 2 所、三级医院 3 所。根据所抽取医院规模大小,采用系统抽样或整群抽样的方法抽取医生,并兼顾性别、年龄和职称等因素。用 SPSS 13.0 软件包对收集的数据进行统计分析。得出医生事务型心理资本得分 4.25 分,人际型心理资本得分 4.41 分;医生心理资本各维度得分最高的是“谦虚沉稳”,得分最低的是“乐观希望”;医生心理资本在性别、年龄、工作总年限、月均收入等方面存在差异,而在职称、医院级别、聘用形式等方面无差异。医生具有较强的心理资本以应对压力。说明增强医生心理资本,对提高医疗服务质量具有重要的实践意义。

【案例辨析与释疑】 在表 1-9 中,不同年龄段医生在人际型心理资本中的“感恩奉献”和“尊敬礼让”两个维度采用方差分析结果有统计学意义,只能说明不同年龄段医生在人际型心理资本中的“感恩奉献”和“尊敬礼让”两个维度的得分不全相等。不能直接得出:年龄在 25 岁及以下的医生,在“感恩奉献”和“尊敬礼让”方面表现最强,而 46 岁及以上的医生两方面表现最差。文章虽有说明两两比较采用 LSD 法和 SNK 法,却未见统计分析结果。

表 1-9 部分人口学变量的医生心理资本差异分析结果 ($\bar{x} \pm s$)

人口学变量		奋 发 图 强	坚 韧 顽 强	乐 观 希 望	自 信 勇 敢	包 容 宽 恕	谦 虚 沉 稳	感 恩 奉 献	尊 敬 礼 让
性别	男	4.28 ± 1.00	4.36 ± 1.10	4.06 ± 0.90	4.44 ± 0.80	4.08 ± 1.00	4.62 ± 0.90	4.50 ± 0.90	4.26 ± 1.00
	女	4.00 ± 0.90	4.32 ± 0.90	4.04 ± 0.90	4.42 ± 0.80	4.16 ± 0.90	4.74 ± 0.80	4.60 ± 0.90	4.50 ± 0.90
	<i>t</i> 值	3.01	0.31	0.25	0.09	-1.18	-1.87	-1.05	-2.76
年龄	≤25 岁	4.56 ± 0.80	4.64 ± 0.80	4.36 ± 1.00	4.64 ± 0.70	4.32 ± 0.90	4.98 ± 0.07	4.78 ± 0.70	4.72 ± 0.80
	26 ~ 35 岁	4.12 ± 1.00	4.28 ± 1.00	4.00 ± 0.90	4.34 ± 0.80	4.06 ± 0.90	4.62 ± 0.80	4.44 ± 0.90	4.32 ± 0.90
	36 ~ 45 岁	4.22 ± 1.00	4.38 ± 1.10	4.06 ± 0.90	4.50 ± 0.70	4.12 ± 1.00	4.72 ± 0.90	4.68 ± 0.90	4.40 ± 0.90
	≥46 岁	4.08 ± 0.90	4.32 ± 1.00	4.08 ± 0.90	4.40 ± 0.90	4.14 ± 0.90	4.56 ± 0.90	4.38 ± 1.00	4.14 ± 0.90
	<i>F</i> 值	2.21	1.20	1.71	2.21	0.82	2.34	3.98	3.37

【案例 1-19】 精神病患者家属对护理服务的满意度调查中,选取杭州市某精神病三级甲等专科医院的出院患者家属为调查对象,采用自行设计的住院患者家属对护理服务满意度调查表进行调查,内容包括基本资料部分和满意度测量部分。护理服务满意度采用描述性统计方法,不同背景家属对护理服务满意度比较采用“检验和”检验。调查资料用 SPSS 11.0 软件建立数据库,进行统计分析处理。得到不同性别、不同职业、不同文化水平、不同收入水平家属对于护理满意度有差别。

【案例辨析与释疑】 本文在对不同职业、不同文化水平、不同收入水平家属对于护理满意度的比较时采用了方差分析,差异具有统计学意义,只能说明不同职业、不同文化水平、不同收入水平家属对于护理满意度不全相同。以表 1-10 中不同文化水平家属对护理满意度的比较为例,方差分析只能说明不同文化水平家属对护理满意度不全相等,而无法得出大专以上文化水平家属对服务礼仪、心理护理要求较高。如果要比较某两组之间是否不同,应该再进行两两比较,并列出现相关统计结果。

表 1-10 不同文化水平家属对护理满意度比较($\bar{x} \pm s$)

文化水平	护理满意度				
	服务礼仪	住院舒适	心理护理	护理技能	总体评价
本科以上($n=39$)	3.97 ± 0.24	4.03 ± 0.34	3.67 ± 0.48	4.11 ± 0.46	4.09 ± 0.32
大专($n=33$)	4.02 ± 0.44	4.08 ± 0.42	3.58 ± 0.49	4.11 ± 0.39	4.02 ± 0.37
高中或中专($n=66$)	4.20 ± 0.40	4.07 ± 0.36	3.90 ± 0.45	4.09 ± 0.36	4.16 ± 0.37
初中或以下($n=58$)	4.24 ± 0.43	4.07 ± 0.44	3.89 ± 0.56	4.11 ± 0.41	4.17 ± 0.35
F 值	5.26	0.12	4.61	0.54	1.52
P 值	0.002	0.949	0.004	0.983	0.211

【案例 1-20】 为了解女性乳腺癌患者信息需求及影响因素,于 2009 年 7 月到 2010 年 12 月采用方便抽样的方法抽取 2 家长沙市三级甲等医院作为调查点,再随机抽取符合条件的患者参与调查,进行现场调查问卷,采用 SPSS 17.0、Amos 7.0 建立数据库并进行统计处理,女性乳腺癌患者的总需求得分为 65.61 ± 20.854 ,且处于手术治疗阶段、年轻、收入高的患者需求强烈。提示不同的社会属性决定信息需求不同。

【案例辨析与释疑】 本文在进行不同家庭月收入的需求比较时采用方差分析,得出不同家庭人均月收入分组患者的需求差异有统计学意义,见表 1-11。而在讨论部分提及收入较高的患者对于信息需求更强烈,该结论无法由方差分析结果得出。如果要比较组与组之间是否不同,不能直接由方差分析得出,应该再进行两两比较,并列出现相关统计结果。

表 1-11 不同家庭人均收入患者需求得分比较

家庭月收入(元)	人 数	均 数	标 准 差
<500	33	55.745	18.177
500~999	73	63.670	18.585
1000~1999	93	64.897	20.359
2000~2999	66	67.551	21.951
3000 以上	58	72.629	22.358

注: $F=3.957$, $P=0.004$ 。

1.3.7 误用 t 检验处理单因素多水平设计定量资料

【案例 1-21】 为了解浙江省流动人口生存质量及影响因素,用分层抽样的方法,依据 WHO 生存质量测定量表(WHO QOL-BREF 量表)测定浙江省 1217 名流动人口和 1387 名常住人口的生存质量,拟合多元线性回归模型分析流动人口生存质量的影响因素。发现流动人口与常住人口在生理领域、社会关系领域和总的健康状况方面差异无统计学意义($P>0.05$),但在心理领域、环境领域得分及总的生存质量低于常住人口(3.5 ± 0.8 VS 3.6 ± 0.8)。影响流动人口心理领域得分的因素为文化程度、家庭摩擦和食欲;影响流动人口环境领域得分的因素为平均每天工作时间和食欲。说明流动人口生存质量虽然已有所改善,但部分领域的得分依然较低,尤其需重点关注心理和环境领域。

【案例辨析与释疑】 本文在进行 WHO QOL-BREF 量表各领域得分情况分析时,将浙江省外来流动人口各领域得分与对照组(常住人口)、国内一般人群和全球 24 个中心平均水平进行比较时,采用两独立样本的 t 检验。表 1-12 中的生理领域、心理领域、社会关系领域、环境领域的比较应采用单因素多水平设计定量资料方差分析,若直接采用 t 检验处理单因素多水平设计定量资料,会增大犯 I 类(医学统计学中常称为假阳性)错误的概率。

表 1-12 流动人口与常住人口、国内一般人群及全球 24 个中心资料生存质量各领域得分比较

人员类别	生理领域	心理领域	社会关系领域	环境领域	总生存质量	总健康状况
流动人口	14.6 ± 2.1	13.8 ± 2.1	14.5 ± 2.4	12.6 ± 2.3	3.5 ± 0.8	3.6 ± 0.8
常住人口	14.5 ± 2.1	14.0 ± 2.2	14.6 ± 2.4	13.5 ± 2.3	3.6 ± 0.8	3.6 ± 0.8
国内一般人群	15.8 ± 2.9	14.3 ± 2.5	13.7 ± 3.0	13.2 ± 2.4	—	—
全球 24 个中心	19.2 ± 2.9	15.0 ± 2.8	14.3 ± 3.2	13.5 ± 2.6	—	—
t_1 值	1.690	2.980	0.510	10.330	4.175	0.441
t_2 值	19.751	8.681	11.722	9.546	—	—
t_3 值	26.376	20.512	3.182	14.006	—	—

注: t_1 流动人口与常住人口比较, t_2 流动人口与国内一般人群比较, t_3 流动人口与全球 24 个中心比较; — 表示无数据。

【案例 1-22】 原文通过对某院单纯胆囊结石手术治疗患者的住院费用的分析,从中找出了影响费用的相关因素,为该院病种纳入单病种定额付费管理提供科学依据。分析不同参保类型采用 t 检验:经检验,三者之间的差别并无明显的统计学意义。其中,在职和退休患者之间 $t = 1.35$, 自由度 = 157, $P = 0.179$; 在职和其他类型患者之间 $t = 0.83$, 自由度 = 104, $P = 0.410$; 退休和其他类型患者之间 $t = 0.21$, 自由度 = 65, $P = 0.833$ 。

【案例辨析与释疑】 该资料中在职患者、退休患者、其他参保类型患者的平均费用比较属于单因素三水平设计定量资料统计分析,首先应该检验资料是否满足参数检验的前提条件,满足则选用单因素三水平设计定量资料方差分析,否则应选用此设计定量资料 Kruskal-Wallis H 检验。差异具有统计学意义时再进行两两比较。可选用的两两比较方法有 q 检验(参数检验)、Nemenyi 法检验(非参数检验)等。

1.3.8 资料对比未考虑重要非试验因素的影响

【案例 1-23】 同【案例 1-21】。

【案例辨析与释疑】 除了误用 t 检验处理单因素多水平设计定量资料外,也未考虑重要非试验因素的影响。文中在一般情况分析时,流动人口的平均年龄低于常住人口的平均年龄,差异有统计学意义($t = 8.162$, $P < 0.05$)。在进行 WHO QOL-BREF 量表各领域得分情况分析时(表 1-12),将浙江省外来流动人口各领域得分与对照组(常住人口)、国内一般人群和全球 24 个中心平均水平进行比较,未考虑年龄这个重要非试验因素可能会对结果造成影响,可按年龄分层,采用分层分析。

1.4 定性资料统计分析中的常见问题及案例辨析

1.4.1 未进行统计分析直接得出结论

【案例 1-24】 不同的药物在用法、用量及服用时间等方面均有可能有差别,细微的差别可能会对治疗结果产生较大的影响。在共同确定用药方案时,临床医师对药物的用药频率等进行指导,促进药物的正确使用。该试验的目的是比较用药频率干预前后的正确率。临床医师重点选择了 3

个科室,对4种药物的用药频率进行干预、指导,并与同期用药频率干预前的结果进行对比,详见表1-13。结果发现其用药频率正确率明显提高。

表 1-13 4 种抗菌药物用药频率干预前后正确率的比较

药 品	用 药 人 次	用药频率正确率(%)	用 药 人 次	用药频率正确率(%)
	干 预 前		干 预 后	
注射用头孢匹胺钠	89	53	142	100
注射用头孢他啶	523	62	446	100
注射用头孢哌酮钠/舒巴坦钠	516	48	610	100
注射用头孢吡肟	92	68	189	100

【案例辨析与释疑】 该资料是通过抽取获得的样本,而原作者未对资料进行统计分析就直按下结论是错误的。该资料是定性资料,可用 χ^2 检验来分析4种抗菌药物用药频率干预前后正确率的差异,注射用头孢匹胺钠: $\chi^2 = 81.90, P < 0.001$;注射用头孢他啶: $\chi^2 = 213.56, P < 0.001$;注射用头孢哌酮钠/舒巴坦钠: $\chi^2 = 415.78, P < 0.001$;注射用头孢吡肟: $\chi^2 = 66.43, P < 0.001$ 。认为4种抗菌药物用药频率正确率干预前后差异有统计学意义,干预后用药频率正确率明显提高。

1.4.2 未标明统计学方法

【案例1-25】 目的是通过对不同举办主体社区卫生服务的房屋来源的比较,了解不同举办类型机构的投入状况。统计学方法未写出,结果得出,在房屋来源方面,政府直接举办和政府所属医疗机构举办的机构状况较好,社会团体或个人举办的机构发展支撑条件较差, $P = 0.00$ 。

表 1-14 不同举办主体社区卫生服务中心房屋来源情况(个)

类 型	免 费 使 用	自 有	租 用	其 他
政府举办	65	166	38	3
公立医院举办	19	22	14	1
企事业单位举办	11	10	8	4
民营	0	2	16	0
合计	95	200	86	8

【案例辨析与释疑】 文中未标出所使用统计学方法而直接得出结论缺乏说服力。根据文中原意是要比较政府直接举办和政府所属医疗机构举办与社会团体和个人举办的机构状况的差异,所以可以将表1-14进行合并见表1-15。

表 1-15 不同举办主体社区卫生服务中心房屋来源情况(个)

类 型	免 费 使 用	自 有	租 用	其 他
政府和政府所属医疗机构举办	84	188	52	4
社会团体和个人举办	11	12	34	4
合计	95	200	86	8

采用 χ^2 检验进行两组间构成比的比较,得 $\chi^2 = 59.52, P < 0.001$,可认为在房屋来源方面政府直接举办和政府所属医疗机构举办与社会团体和个人举办的差异有统计学意义。

【案例1-26】 目的是分析三位一体行为干预的效果。设立试验组 and 对照组,对A组进行化疗+三位一体行为干预,B组只做化疗,观察两组的不良反应与生活质量的差异详见表1-16。未说明

用何种统计学方法。结果认为两组不良反应比较差异均有统计学意义，两组患者生活质量比较差异均有统计学意义。

表 1-16 两组患者不良反应比较(例)

组 别	例 数	食 欲 降 低	恶 心	呕 吐	乏 力
A 组	46	16	10	6	12
B 组	4627	20	15	28	
P 值		< 0.05	< 0.05	< 0.05	< 0.01

表 1-17 两组患者生活质量比较(KPS)

组 别	例 数	显 著 差 异	改 善	稳 定	下 降	改善率(%)
A 组	46	12	18	11	5	65.2
B 组	46	6	12	8	18	39.1

【案例辨析与释疑】 原作者要对 A 组与 B 组间的差别进行比较，分别对每种不良反应进行 χ^2 检验，食欲降低： $\chi^2 = 5.28, P = 0.022$ ；恶心： $\chi^2 = 4.95, P = 0.026$ ；呕吐： $\chi^2 = 4.99, P = 0.025$ ；乏力： $\chi^2 = 11.32, P = 0.001$ ；认为两组不良反应发生率差异有统计学意义。表 1-17 为单向有序资料，宜采用秩和检验，得 $Z = 2.872, P = 0.004$ ，认为两组患者的生活质量差异有统计学意义。

1.4.3 误用 χ^2 检验分析有序列联表资料

【案例 1-27】 目的是了解流动人口与当地常住人口基本公共卫生服务不均等的情况。采用 χ^2 检验分析流动人口与当地常住人口的年龄差异，结果 $\chi^2 = 108.348, P = 0.000$ ，得出两组之间年龄差异有统计学意义的结论。

表 1-18 调查人群的社会学特征

年 龄	人 数	百分比(%)	人 数	百分比(%)	χ^2	P
	流 动 人 口		常 住 人 口			
15 ~ 29	300	51.7	167	28.9		
30 ~ 44	242	41.7	256	44.4	108.343	0.000
45 以上	38	6.6	154	26.7		

【案例辨析与释疑】 表 1-18 中资料为单向有序资料，采用秩和检验分析资料得， $Z = 9.904, P < 0.001$ ，认为流动人口与常住人口的年龄差异有统计学意义，常住人口年龄大于流动人口。

【案例 1-28】 目的是探讨高血压患者对医疗保险的满意度，以提高医疗保险经办机构的管理水平和管理质量。采用 χ^2 检验分析不同地区高血压患者对医疗保险满意度的差异。结果 $\chi^2 = 259.535, P < 0.05$ ，认为由于各地医疗保险政策的不同，各地调查对象对保险的评价也不同。

表 1-19 高血压患者对医疗保险满意度评价

满 意 度	人 次	百分比(%)	人 次	百分比(%)	人 次	百分比(%)	人 次	百分比(%)
	北京东城区		南京秦淮区		杭州下城区		合 计	
非常满意	20	3.4	27	4.6	107	19.2	154	8.9
比较满意	190	32.7	152	26.1	229	41.1	571	33.2
一般	132	22.7	138	23.7	142	25.5	412	23.9
不太满意	129	22.3	201	34.5	53	9.5	383	22.3
很不满意	39	6.7	38	6.5	5	0.9	82	4.8
不清楚	71	122.2	27	4.6	21	3.8	119	6.9

【案例辨析与释疑】表 1-19 资料为单向有序资料,故不宜应用 χ^2 检验,应用秩和检验。且原表中满意度的排列顺序将“不清楚”排最后,不符合一般的满意度递减顺序,故将表 1-19 中满意度重新排列如表 1-20。

表 1-20 高血压患者对医疗保险满意度评价

满 意 度	人 次	百分比(%)	人 次	百分比(%)	人 次	百分比(%)	人 次	百分比(%)
	北京东城区		南京秦淮区		杭州下城区		合 计	
非常满意	20	3.4	27	4.6	107	19.2	154	8.9
比较满意	190	32.7	152	26.1	229	41.1	571	33.2
一般	132	22.7	138	23.7	142	25.5	412	23.9
不清楚	71	12.2	27	4.6	21	3.8	119	6.9
不太满意	129	22.3	201	34.5	53	9.5	383	22.3
很不满意	39	6.7	38	6.5	5	0.9	82	4.8

用秩和检验分析表 1-20 资料得 $\chi^2 = 192.914$, $P < 0.001$,认为北京东城区、南京秦淮区、杭州下城区三地高血压患者对医疗保险满意度的差异有统计学意义。

1.4.4 误用 t 检验处理定性资料

【案例 1-29】抽取实施临床途径管理急性化脓性阑尾炎患者 142 例作为试验组,同时在两家样本医院分别抽取未实施临床路径管理同期同类的阑尾炎患者 116 例作为对照组,抽取试验组和对照组患者的病案首页,见表 1-21。目的是分析试验组和对照组在性别和年龄方面的差异。采用 t 检验分析。得出结果,认为试验组和对照组在性别和年龄方面均无显著性差异。

表 1-21 样本医院病案首页抽取情况

分 项		例数(N)	百分比(%)	例数(N)	百分比(%)	t 值	P 值
性别	男	85	59.9	60	51.7	-1.309	0.193
	女	57	40.1	56	48.3		
年龄	<20	22	15.5	18	15.5		
	20~39	85	59.9	68	58.6		
	40~59	27	19.0	12	10.3	-1.058	0.291
	≥60	8	5.6	18	15.5		
	合计	142		116			

【案例辨析与释疑】表 1-21 中性别与年龄分组均为定性资料,而 t 检验是用于定量资料的分析方法,所以原作者用分析定量资料的方法分析定性资料是错误的。对于性别可以用 χ^2 检验分析试验组与对照组之间的差异,得 $\chi^2 = 1.716$, $P = 0.190$,尚不能认为试验组和对照组的性别构成差异有统计学意义。因为年龄可被视为单向有序的结果变量,所以用秩和检验分析两组的年龄差异,得 $Z = 0.527$, $P = 0.598$,尚不能认为试验组和对照组的年龄差异有统计学意义。

1.4.5 误用 χ^2 检验回答一致性问题

【案例 1-30】目的是分别调查患者家属对告知患者真实病情的态度、患者了解自己病情的意愿,以核对二者是否一致,从而考察患者的知情权是否得到了应有的尊重。采用 χ^2 分析患者不同知情状态下家属与患者态度的一致性的差异。结果发现患者是否知情显著影响患者与家属间的一致性,患者知晓病情与患者被隐瞒病情两组间的患者与家属的一致性差异有统计学意义($P = 0.000$),见表 1-22 和表 1-23。

表 1-22 家属的态度与患者的知情意愿比较

家属与患者的态度	例 数	百分比 (%)	例 数	百分比 (%)
	家属的态度		患者的知情意愿	
告知/想知道病情	53	52.5	86	85.5
瞒/不想知道病情	48	47.5	15	13.5

表 1-23 患者不知情状态下家属与患者态度一致性比较

家属与患者的态度是否一致	例 数	百分比 (%)	例 数	百分比 (%)
	患者知晓病情 (n=48)		患者被隐瞒病情 (n=23)	
一致	36	75.0	7	30.0
不一致	12	25.0	16	70.0

【案例辨析与释疑】 原作者目的是要分析患者家属对告知患者真实病情的态度与患者了解自己病情的意愿的一致性，而文中未提到二者间的一致性如何，只是将表 1-23 用 χ^2 检验分析了患者是否知晓病情会不会影响家属与患者的一致性。根据原文献内容“在 101 对恶性肿瘤患者及其家属中，有 61.4% (62/101) 的家属态度与患者的意愿是一致的”及表 1-22 编制配对四格表。结合一致性系数 Kappa 值评定标准(注意：此标准对总体而言)：小于 0.4 为一致性弱，在 0.4 ~ 0.75 之间为一致性中等，大于 0.75 为一致性强，给出患者与家属态度的一致性。

表 1-24 家属态度与患者意愿一致性

家 属 态 度	人 数		
	患者意愿： 想知道	不 想 知 道	合 计
告知	50	3	53
隐瞒	36	12	48
合计	86	15	101

分析表 1-24 得 Kappa 值为 0.20，小于 0.4， $P=0.073$ ，可认为一致性弱且无统计学意义。

1.4.6 误用一般 χ^2 检验说明趋势问题

【案例 1-31】 目的是分析不同干预措施下医院总的吸烟有效劝阻率、室内与室外吸烟有效劝阻率、垃圾桶旁与其他区域吸烟有效劝阻率差异。采用一般 χ^2 分析。结果认为不同干预措施下医院总的吸烟有效劝阻率呈递增趋势($P < 0.05$ ，差异有统计学意义)，对室内吸烟的有效劝阻率呈持续上升趋势($P < 0.05$ ，差异有统计学意义)，对室外吸烟有效劝阻率差异无统计学意义($P > 0.05$)，垃圾桶旁的有效劝阻率虽有小幅上升但差异无统计学意义($P > 0.05$)，非垃圾桶旁的有效劝阻率却随着控烟措施的逐步推进，呈明显上升趋势($P < 0.05$ ，差异有统计学意义)，详见表 1-25 ~ 表 1-27。

表 1-25 不同干预措施下吸烟总的有效劝阻率

调 查 时 间	人 数		合计人数	有效劝阻率 (%)
	劝阻效果： 有效	无 效		
第一阶段实施后	55	89	144	38.2
第二阶段实施后	43	53	96	44.8
第三阶段实施后	60	29	89	67.4

注： $\chi^2 = 19.384$ ， $P = 0.000$ 。

表 1-26 不同干预措施下室内与室外吸烟有效劝阻率的比较

调查时间	有效人数	无效人数	有效劝阻率(%)	有效人数	无效人数	有效劝阻率(%)
	劝阻室内吸烟情况			劝阻室外吸烟情况		
第一阶段实施后	19	52	26.8	36	37	49.3
第二阶段实施后	26	32	44.8	17	21	44.7
第三阶段实施后	25	10	71.4 *	35	19 *	64.8 ㉔

注: * $\chi^2 = 19.288$, $P = 0.000$, ㉔ $\chi^2 = 4.462$, $P = 0.107$ 。

表 1-27 垃圾桶旁与其他区域吸烟有效劝阻率的比较

调查时间	有效人数	无效人数	有效劝阻率(%)	有效人数	无效人数	有效劝阻率(%)
	在垃圾桶旁的劝阻情况			非垃圾桶旁的劝阻情况		
第一阶段实施后	14	18	43.8	41	71	36.6
第二阶段实施后	7	8	46.7	36	45	44.4
第三阶段实施后	1	1	50.0 *	59	28	57.8 ㉔

注: * $\chi^2 = 0.057$, $P = 0.972$; ㉔ $\chi^2 = 19.869$, $P < 0.001$ 。

【案例辨析与释疑】一般 χ^2 检验只能说明各行上有效率间的差异是否具有统计学意义,并不能就此说明趋势问题。原作者未对资料进行假设检验,直接根据数据观察得出结论是不可取的,也是不可信的。可以应用线性趋势 χ^2 检验分析其升降趋势。经计算不同干预措施下医院总的吸烟有效劝阻率 $\chi^2 = 19.384$, $P < 0.001$,认为不同干预措施下医院总的吸烟有效劝阻率差异有统计学意义;趋势 $\chi^2 = 17.611$, $P < 0.001$,认为不同干预措施下医院总的吸烟有效劝阻率呈递增趋势。对室内吸烟的有效劝阻率 $\chi^2 = 19.288$, $P < 0.001$,认为不同干预措施下对室内吸烟的有效劝阻率差异有统计学意义;趋势 $\chi^2 = 18.905$, $P < 0.001$,认为不同干预措施下对室内吸烟的有效劝阻率呈持续上升趋势。对室外吸烟有效劝阻率 $\chi^2 = 4.462$, $P = 0.107$,认为不同干预措施下对室外吸烟的有效劝阻率差异无统计学意义。垃圾桶旁的有效劝阻率 $\chi^2 = 0.057$, $P = 0.972$,认为不同干预措施下垃圾桶旁的有效劝阻率差异无统计学意义。非垃圾桶旁的有效劝阻率 $\chi^2 = 19.869$, $P < 0.001$,认为不同干预措施下非垃圾桶旁的有效劝阻率差异有统计学意义;趋势 $\chi^2 = 18.419$, $P < 0.001$,认为不同干预措施下非垃圾桶旁的有效劝阻率呈递增趋势。

1.5 相关与回归分析中的常见问题及案例辨析

1.5.1 不同相关分析方法应用条件上混淆不清

【案例 1-32】目的是评价医院医疗工作绩效综合评价值与月收支结余存在的关系。采用皮尔逊相关系数与秩相关系数分析非手术科室、手术科室的医疗工作绩效综合评价值与月收支结余存在的关系,详见表 1-28 ~ 表 1-31。结果认为非手术科室的综合评价值与各科室月收支结余间的相关系数均为正值,皮尔逊相关系数为 0.570,有统计学意义;手术的综合评价值与各科室月收支结余的相关系数亦均正值,秩相关系数为 0.566,有统计学意义。

表 1-28 非手术科室综合评价值(C值)与收支结余排名

科室名称	C值(降序)	收支结余排名	科室名称	C值(降序)	收支结余排名
A2	0.646	2	A4	0.332	10
A7	0.575	9	A9	0.312	7
A8	0.506	1	A5	0.279	11
A1	0.480	4	A11	0.224	5

续表

科 室 名 称	C 值(降序)	收支结余排名	科 室 名 称	C 值(降序)	收支结余排名
A6	0.453	3	A10	0.143	6
A3	0.397	8			

表 1-29 手术科室综合评价值(C 值) 与收支结余排名

科 室 名 称	C 值(降序)	收支结余排名	科 室 名 称	C 值(降序)	收支结余排名
B6	0.615	2	B5	0.282	3
B9	0.536	4	B7	0.248	6
B12	0.464	9	B10	0.239	12
B8	0.331	1	B1	0.221	10
B3	0.321	7	B2	0.216	5
B11	0.308	8	B4	0.129	11

表 1-30 非手术科室综合评价值(C 值) 与收支结余的相关分析结果

变 量	C 值		收 支 结 余		C 值		收 支 结 余	
	皮尔逊相关系数	单侧 P 值	皮尔逊相关系数	单侧 P 值	秩相关系数	单侧 P 值	秩相关系数	单侧 P 值
C 值	1		0.570	0.034	1		0.427	0.095
收支结余	0.570	0.034	1		0.427	0.095	1	

表 1-31 手术科室综合评价值(C 值) 与收支结余的相关分析结果

变 量	C 值		收 支 结 余		C 值		收 支 结 余	
	皮尔逊相关系数	单侧 P 值	皮尔逊相关系数	单侧 P 值	秩相关系数	单侧 P 值	秩相关系数	单侧 P 值
C 值	1		0.287	0.183	1		0.566	0.027
收支结余	0.287	0.183	1		0.566	0.027	1	

【案例辨析与释疑】 Pearson 相关分析适用于双变量正态分布资料，而 Spearman 秩相关分析适用于不服从双变量正态分布或数据为等级资料或者分布类型不清的资料。而在表 1-28、表 1-29 所示资料中原作者将二者都用于分析同一资料，只选取有统计学意义的结果来陈述结论是不正确的。该数据应采用 Spearman 秩相关分析，只需报告秩相关分析结果，包括 Spearman 秩相关系数及其假设检验的具体 P 值。

1.5.2 过小的 r 值在下肯定相关结论时所遇到的问题

【案例 1-33】 目的是进行单纯胆囊结石手术治疗患者住院费用的相关因素分析。采用相关分析(文中未指明是哪种相关)。结果认为性别、手术方式与住院费用呈现强相关，患者日均费用、住院天数、自付费用也与总体费用水平呈现出强相关性，见表 1-32。

表 1-32 单纯胆囊结石手术治疗患者住院费用的相关因素分析

相 关 因 素	相 关 系 数	P 值	相 关 因 素	相 关 系 数	P 值
性别	-0.154	0.048	放射费	0.383	0.000
日均费用	-0.217	0.005	化验费	0.452	0.000
住院天数	0.665	0.000	输氧费	0.339	0.000
床位费	0.441	0.000	诊疗费	0.65	0.000
药费	0.822	0.000	麻醉费	0.356	0.000
西药费	0.816	0.000	陪床费	0.503	0.000
检查费	0.453	0.000	自付费用	0.526	0.000
治疗费	0.481	0.000	手术方式	-0.155	0.046

【案例辨析与释疑】 相关系数 $|r| < 0.4$ 为弱相关, $0.4 < |r| < 0.7$ 为中度相关, $|r| > 0.7$ 为强相关(注意:此标准是对总体而言的)。表1-32中的日均费用与住院费用的相关为弱相关,药费、西药费与住院费用的相关为强相关,其他因素与住院费用的相关为中度相关。当然,仅当相关分析所得的 r^2 或 r_s^2 大于0.5且 $P < 0.05$ 时,可能更有实用价值。另外,表中的性别、手术方式这两个变量是分类变量,与患者住院费用的相关分析,不宜直接计算相关系数,须改用其他分析方法。

1.5.3 用单因素分析取代多因素分析

【案例1-34】 目的是调查临床护士工作满意度并分析其影响因素。采用 t 检验和方差分析处理满意度相关因素。结果认为不同科室、不同职称、不同月收入水平、不同年龄、不同工龄、不同从事护理的目的都可能导致护士的工作满意度差异,见表1-33。

表1-33 护士工作的总体满意度比较

项 目		人 数	百分比 (%)	满意度平均值	标 准 差	F 值	P 值
职称	护 士	87	34.9	3.41	0.56	4.727	0.010
	护 师	124	49.8	3.32	0.57		
	主管护师以上	38	15.3	3.65	0.60		
月收入	< 2000	59	23.7	3.26	0.47	3.936	0.009
	2000 ~ 2999	127	51	3.38	0.59		
	3000 ~ 3999	56	22.5	3.53	0.62		
	> 4000	7	2.8	3.88	0.50		
工龄	< 10	132	53.0	3.28	0.58	5.024	0.002
	11 ~ 20	90	36.2	3.34	0.57		
	21 ~ 30	22	8.8	3.57	0.51		
	> 30	5	2.0	4.27	0.14		

注:表格内容过多,在此只截取部分数据。

【案例辨析与释疑】 本文分析属于多因素分析,原作者在分析护士工作满意度的影响因素时,只分析各个因素内部差别有无差异,并不能说明该因素对护士满意度的影响有统计学意义。因为单独分析该因素与护士工作满意度的关系,并不能消除其他因素的混杂作用。本例资料只进行了单因素分析,应在此基础上进一步采用多因素分析方法,如多重线性回归分析,采用逐步回归法、前进法和后退法分别进行变量筛选(以统一的标准选择其中一种最恰当的分析结果),得到有统计学意义的变量,即对护士工作满意度有意义的影响因素。

1.6 高级统计分析方法中的常见问题及案例辨析

1.6.1 主成分分析——指标选择条件不满足

【案例1-35】 用文献法和多种统计学方法建立指标体系,运用主成分分析法和TOPSIS法进行统计分析,确定涵盖科室社会效益、医疗工作总量、人均工作量、工作效率和医疗工作质量5个方面(一级指标)的共26个(非手术科室)或32个(手术科室)二级指标用于评价临床科室医疗工作绩效。

【案例辨析】 在获得各一级指标综合评价值,分别对5个一级指标所属的二级指标进行主成分法分析发现,非手术科室内只有医疗工作总量指标(KMO为0.589, $P < 0.001$)和人均工作量指

标(KMO 为 0.684, $P < 0.001$)适宜用该法获得综合评价价值;手术科室内仅医疗工作总量指标满足主成分分析的条件(KMO 为 0.582, $P < 0.001$)。在指标的选择时 KMO 均小于 0.7,可能不适合做主成分分析。应再采取其他方法进行探讨,综合得出相应结论。

【案例释疑】 KMO(Kaiser-Meyer-Olkin)检验统计量是用于比较变量间简单相关系数和偏相关系数的指标,主要应用于多元统计的主成分分析和因子分析。KMO 统计量取值在 0 和 1 之间。当所有变量间的简单相关系数平方和远大于偏相关系数平方和时,KMO 值接近 1。KMO 值越接近 1 意味着变量间的相关性越强,原有变量越适合作主成分分析和因子分析;当所有变量间的简单相关系数平方和接近 0 时,KMO 值接近 0。KMO 值越接近 0 意味着变量间的相关性越弱,原有变量越不适合做主成分分析和因子分析。Kaiser 给出了常用的 KMO 度量标准:0.9 以上表示非常适合;0.8 表示适合;0.7 表示一般;0.6 表示不太适合;0.5 以下表示极不适合。

1.6.2 重测信度——评价指标有误

【案例 1-36】 通过对护士长核心能力评估体系的探索,挑选出代表性好、有一定区别能力、又较少关联的指标组成评估体系,为护士长核心能力的评估提供依据。通过评估指标筛选方法: t 检验法、反应度分析、相关系数法、因子分析法、Cronbach α 系数、重测信度法。结合统计学理论分析,在以上标准中达到任意 3 条,并综合考虑临床工作中护士长的能力水平,删除 8 个指标,最终形成的“护士长核心能力评估体系”共 6 个维度 33 个指标。以 Cronbach α 系数考察同质性信度, Pearson 相关系数考察重测信度。通过专家咨询进行内容效度检测,以内容效度指数(Content Validity Index, CVI)、内部一致性和因子分析评估体系的效度。经过 KMO 值和 Bartlett 球形检验说明适合进行因子分析。

【案例辨析与释疑】 重测信度评价方法:若评估指标为定量数据,一般认为采用组内相关系数(Intra-class Correlation Coefficient, ICC)比采用 Pearson 相关系数好一些;若评估指标为等级数据,应采用 Kappa 系数。值得注意的是,重测信度本质上是一致性评价问题,而相关系数对系统误差很不敏感。因此,更为理想的评价重测信度的方法尚待进一步研究。

1.6.3 因子分析——公共因子个数选择

【案例 1-37】 在杭州市区社区卫生服务中心的就诊患者中开展问卷调查,就患者对社区医生信任的影响因素进行了因子分析、回归分析,并在此基础上提出改善社区医生信任度水平的措施。

【案例辨析与释疑】 采用主成分分析法,并且通过方差最大法进行正交旋转,提取特征值大于 1 的 4 个公共因子,分别为贴心服务、良好态度、患者利益和经验判断,各解释总体方差的比例(贡献率)为 31.84%、13.46%、10.26% 和 9.74%,4 个公共因子的累积贡献率为 65.3%。公共因子个数选择一般应根据特征值大于 1,公共因子的累积贡献率达到 70% 以上,并且结合各公共因子的实际含义来确定。本案例所选因子累积贡献率小于 70%,应适当增加因子个数,以便保留原始指标的主要信息,得出更有实际意义的结论。

1.6.4 结构方程模型——相关系数大于 1

【案例 1-38】 借鉴国外早已开发并一直广泛使用的通用量表,综合考虑工作满意度构成和影响因素,应用结构方程模型分析探讨影响社区医务人员工作的内在满意度和外在满意度,为提高社区医务人员工作满意度,探索有效干预措施提供依据。

【案例辨析与释疑】 医务人员工作满意度分析时,指出:最终模型的标准化结果显示,在

考虑了工作类型、性别、年龄、教育程度和收入之后,内在满意度和外在满意度之间相关系数为 1.01。

结构方程模型的标准化解有两种形式:完全标准化解和标准化解。习惯上,研究结果报告完全标准化解,此时系数值限定于 -1 至 1 之间,使研究结果易于理解。

1.7 统计结果报告中的常见问题及案例辨析

常见的统计资料的表达与描述方法有统计表和统计图。统计表可以用来反映精确的统计数据,统计图则给人更直观的印象。在统计结果表达中,常可看见一些不恰当甚至错误的表达:如统计表中的横、纵标目指代不明、以比代率、表中数值暂缺用汉字表示,统计表的选用错误,统计图的选用错误、单位未标注等问题。近年来,随着高级统计分析方法(如因子分析、主成分分析、路径分析等)的运用增加,在高级统计学方法的运用中也出现了相应的问题,如选用条件模糊不清等问题。值得注意的是,本节中有些统计表还可进一步修改。

1.7.1 统计表使用的错误辨析与释疑

1. 标题对数值的指代错误——以比代率

【案例 1-39】 部分数据表达见表 1-34。

表 1-34 2004 年和 2008 年虹口区基层医疗机构开展室内质控和参加室间质控情况[例数, 百分比(%)]

	医院数量	室内质控例数(%)				室间质控例数(%)			
		血液	尿液	生化	肝炎标志物	血液	尿液	生化	肝炎标志物
2004 年	20	13(65)	12(60)	12(60)	7(35)	10(50)	10(50)	9(45)	7(35)
2008 年	20	20(100)	20(100)	15(75)	12(60)	20(100)	20(100)	14(70)	11(55)
χ^2 值		8.49	10.00	1.03	2.51	13.33	13.33	2.56	1.62
P 值		0.004	0.002	0.311	0.113	0.000	0.000	0.110	0.204

【案例辨析与释疑】 在表 1-34 的标题中,错用百分比代替百分率。百分率可以用来说明事物或现象发生的强度或频率,百分比常用以说明事物内部各组分所占的比重,因此表 1-34 的标题应改为百分率,见表 1-35。

表 1-35 2004 年和 2008 年虹口区基层医疗机构开展室内质控和参加室间质控情况

	医院数量	室内质控[例数, 百分率(%)]				室间质控[例数, 百分率(%)]			
		血液	尿液	生化	肝炎标志物	血液	尿液	生化	肝炎标志物
2004 年	20	13(65)	12(60)	12(60)	7(35)	10(50)	10(50)	9(45)	7(35)
2008 年	20	20(100)	20(100)	15(75)	12(60)	20(100)	20(100)	14(70)	11(55)
χ^2 值		8.49	10.00	1.03	2.51	13.33	13.33	2.56	1.62
P 值		0.004	0.002	0.311	0.113	0.000	0.000	0.110	0.204

2. 横标目指代模糊不清

【案例 1-40】 部分数据表达见表 1-36。

【案例辨析与释疑】 在表 1-36 中,横标目“2010 年较 2007 年增减(%)”指代有误。采用“增减”指代时数值应该为绝对值,不应有负数出现,而此行的数据有负值,因而结合文章大意,应该采用“2010 年较 2007 年增加(%)”代替,见表 1-37。

表 1-36 2007 年和 2010 年新农合门诊次均结报费用

类 别	门诊次均结报					住院次均结报费用			
	平 均	村	镇 级	区 级	市 级	平 均	镇 级	区 级	市 级
2007 年次均费用(元)	30.9	15.9	51.8	49.3	56.8	2149.3	1129.2	2222.3	3076.4
2010 年次均费用(元)	33.1	18.2	43.4	56.7	67.3	3183.3	1541.3	3396.7	4299.5
2010 年较 2007 年增减(%)	7.3	14.5	-16.3	14.9	18.4	48.1	36.5	52.8	39.8
2007—2010 年均递增(%)	2.4	4.6	-5.8	4.7	5.8	14	10.9	15.2	11.8

表 1-37 2007 年和 2010 年新农合门诊次均结报费用

类 别	门诊次均结报					住院次均结报费用			
	平 均	村	镇 级	区 级	市 级	平 均	镇 级	区 级	市 级
2007 年次均费用(元)	30.9	15.9	51.8	49.3	56.8	2149.3	1129.2	2222.3	3076.4
2010 年次均费用(元)	33.1	18.2	43.4	56.7	67.3	3183.3	1541.3	3396.7	4299.5
2010 年较 2007 年增加(%)	7.3	14.5	-16.3	14.9	18.4	48.1	36.5	52.8	39.8
2007—2010 年均递增(%)	2.4	4.6	-5.8	4.7	5.8	14	10.9	15.2	11.8

3. 纵标目指代模糊不清

【案例 1-41】 部分数据表达见表 1-38。

表 1-38 不同就诊医疗机构的时间成本分析

医 疗 机 构	步行 时 间				就诊等候时间				医生接诊时间			
	<15 分钟		>15 分钟		<60 分钟		>60 分钟		<5 分钟		>5 分钟	
	人次	%	人次	%	人次	%	人次	%	人次	%	人次	%
社康中心	112	79	29	21	134	92	12	8	29	20	117	80
街道医院	10	16	51	84	37	57	28	43	10	15	55	85
区/市医院	20	22	73	78	56	51	53	49	24	22	86	78
民营机构	30	65	16	35	56	88	8	13	17	27	47	73

【案例辨析与释疑】 在表 1-38 中，纵标目的指代不明确，没有对“%”的意义加以说明，不能看出“%”所指的是百分比还是百分率，因此结合文章大意，改为表 1-39。

表 1-39 不同就诊医疗机构的时间成本分析

医 疗 机 构	步行 时 间				就诊等候时间				医生接诊时间			
	<15 分钟		>15 分钟		<60 分钟		>60 分钟		<5 分钟		>5 分钟	
	人次	百分比 (%)	人次	百分比 (%)	人次	百分比 (%)	人次	百分比 (%)	人次	百分比 (%)	人次	百分比 (%)
社康中心	112	79	29	21	134	92	12	8	29	20	117	80
街道医院	10	16	51	84	37	57	28	43	10	15	55	85
区/市医院	20	22	73	78	56	51	53	49	24	22	86	78
民营机构	30	65	16	35	56	88	8	13	17	27	47	73

4. 数值区表达形式错误

【案例 1-42】 部分数据表达见表 1-40。

表 1-40 2007 年和 2008 年城市医院与社区卫生服务机构诊疗服务量对照

年 份	诊疗人次(亿)		入院人数(万)	
	城 市 医 院	社区卫生服务机构	城 市 医 院	社区卫生服务机构
2007 年	10.98	2.26	3586	107
2008 年	11.3	暂缺	3670	暂缺

【案例辨析与释疑】 在表 1-40 中，2008 年的社区服务机构诊疗人次与入院人数的表述“暂缺”应该用“…”代替，见表 1-41。

表 1-41 2007 年和 2008 年城市医院与社区卫生服务机构诊疗服务量对照

年 份	诊疗人次(亿)		入院人数(万)	
	城 市 医 院	社区卫生服务机构	城 市 医 院	社区卫生服务机构
2007 年	10.98	2.26	3586	107
2008 年	11.3	…	3670	…

5. 统计描述无法得出统计推断部分的内容

【案例 1-43】 部分数据表达见表 1-42。

表 1-42 安徽省 8 所公立医院医务人员对改革政策知晓情况

内 容	选 择 项 目	人 数	百分比 (%)	统计学分析(χ^2)				
				医院级别	从事岗位	职 务	职 称	年 龄
政策 A	非常不了解	60	5.3	2.999	16.149 **	22.015 **	19.271 **	11.835
	了解	892	78.2					
	不了解	188	16.5					
政策 B	非常不了解	53	4.6	5.355	9.867 *	19.099 **	10.943 *	9.715
	了解	836	73.3					
	不了解	251	22.0					

注：* 表示 $P < 0.05$ ，** 表示 $P < 0.01$ 。

【案例辨析与释疑】 在表 1-42 中，统计表的绘制有误，表中的“统计学分析(χ^2)”无法由表中数据得出。另外，因为结果变量为有序变量：非常了解、了解、不了解，用一般的卡方检验欠妥，宜改用秩和检验。现以不同医院级别对政策 A 的知晓情况进行统计学分析为例，绘制表 1-43，表中的 u 值为两组比较秩和检验的结果。

表 1-43 安徽省 8 所公立医院不同级别医院的医务人员对改革政策 A 知晓情况

医院 级 别	非 常 了 解	了 解	不 了 解	u 值	P 值
三级医院	a	b	c	g	h
二级医院	d	e	f		

1.7.2 统计图使用的错误辨析与释疑

1. 条图纵轴指代不明

【案例 1-44】 部分数据表达见图 1-1。

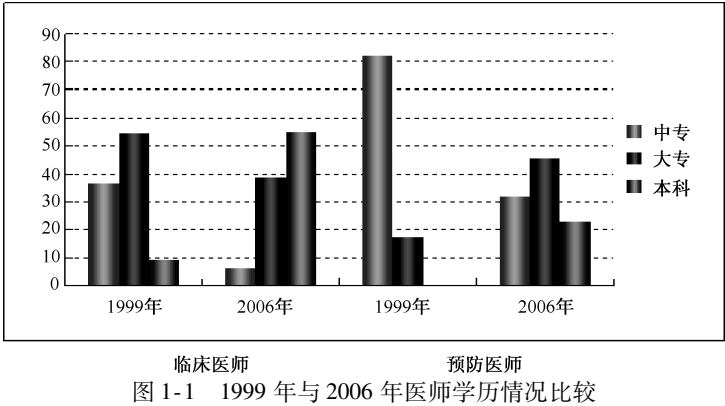


图 1-1 1999 年与 2006 年医师学历情况比较

【案例辨析与释疑】 在图 1-1 中，纵轴数值的指代不明确，结合文章大意，改为图 1-2。

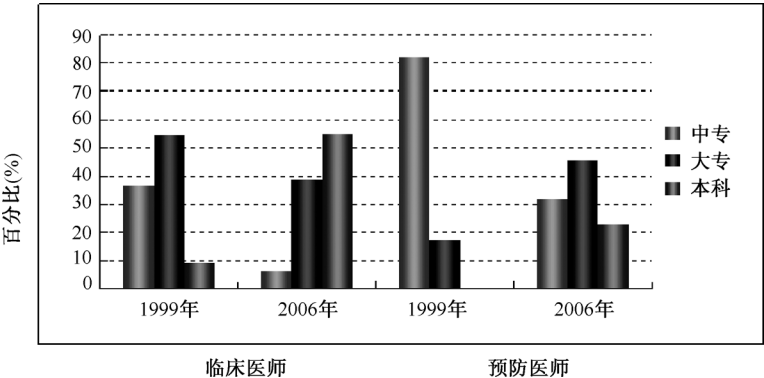


图 1-2 1999 年与 2006 年医师学历情况比较

【案例 1-45】 部分数据表达见图 1-3。

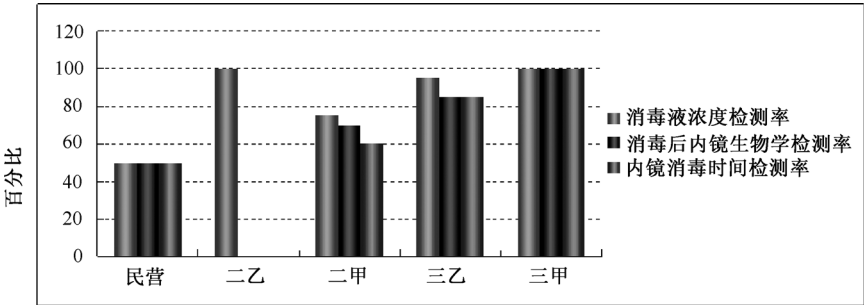


图 1-3 浙江省医院的内镜自检情况

【案例辨析与释疑】 在图 1-3 中，纵轴未写明单位，结合文章大意，改为图 1-4。

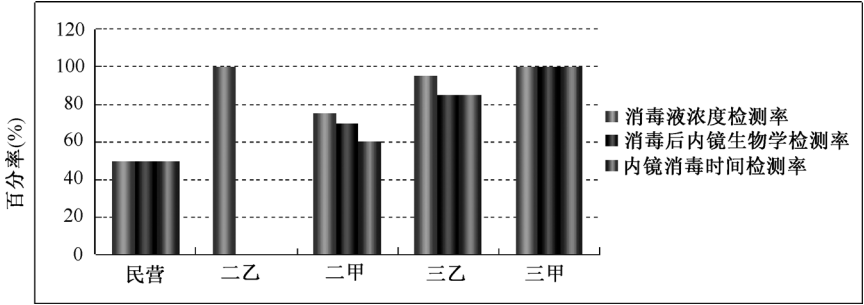


图 1-4 浙江省医院的内镜自检情况

2. 统计图选择错误——误用线图指代条图内容

【案例 1-46】 部分数据表达见图 1-5。

【案例辨析与释疑】 在图 1-5 中，比较境内与境外患者诊疗环境满意率，应该采用条图。线图常常用线段的升降表示统计指标的变化趋势，或一种现象随另一现象的变化情况；而条图常常用以表示相互独立的指标之间的数量的大小，因此结合文章大意，选用条图来表示更为妥当，见图 1-6。

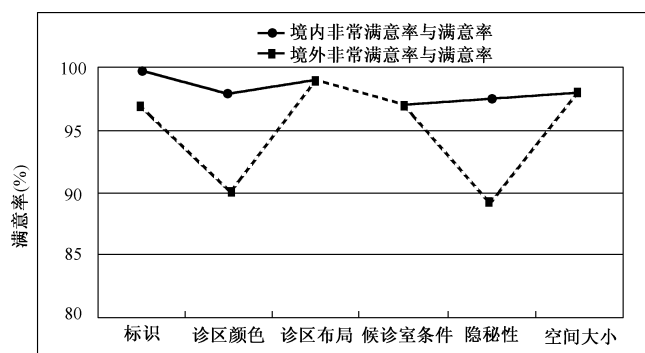


图 1-5 境内与境外患者诊疗环境满意率比较

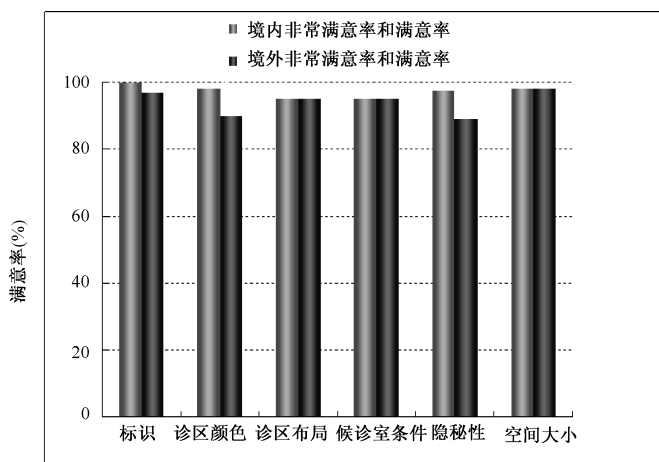


图 1-6 境内与境外患者诊疗环境满意率比较

1.8 本章内容小结

医学科研成就绝大多数通过发表学术论文的方式表现出来。本章重点研究《中华医院管理杂志》在研究设计、统计分析和结果表达方面存在的问题,通过 2007—2011 年发表论文中找出的 47 个典型案例,旨在提高该医学期刊论文的质量,进一步有效地提高医学科研工作的科学性和严谨性。

参考文献

- [1] 胡良平, 李子建. 医学统计学基础与典型错误辨析. 北京: 军事医学科学出版社, 2003.
- [2] 胡良平. 正确实施科研设计与统计分析. 北京: 人民军医出版社, 2011.
- [3] 胡良平. 检验医学科研设计与统计分析. 北京: 人民军医出版社, 2004.
- [4] 胡良平. 医学统计应用错误的诊断与释疑. 北京: 军事医学科学出版社, 1999.
- [5] 刘桂芬. 医学统计学(第 2 版). 北京: 中国协和医科大学出版社, 2007.
- [6] 方积乾. 卫生统计学(第 6 版). 北京: 人民卫生出版社, 2008.
- [7] 胡良平. 统计学三型理论在试验设计中的应用. 北京: 人民军医出版社, 2006.
- [8] Walter SD. Methods of reporting statistical results from medical research studies. Am J Epidemiol, 1995, 141 (10): 896-906.

第2章 危重病急救医学统计设计与数据分析

《中国危重病急救医学》杂志是急救医学领域具有较大影响力的学术期刊，主要发表与危重病急救有关的科研论文，包括临床试验研究、观察性研究和试验研究三种类型，其中又以临床观察性研究和临床试验类研究居多。该杂志对拟刊文章有明确的统计学要求，如必须按照国家标准正确使用统计学符号、应该说明研究类型的名称、正确选择统计分析方法及统计结果的正确表达等。

本章简要介绍危重病急救医学研究的概况，重点结合近几年《中国危重病急救医学》杂志发表文章，按照临床试验研究、观察性研究和试验研究对文章进行分类，每类研究选择若干篇具有代表性的文章，并对每一篇文章从统计设计、数据的统计学处理、统计结果表达等方面进行辨析，以便读者能够全面理解一项科学研究应该考虑的统计学问题，从而更好地服务于医学科研实践。

2.1 急救医学研究概述

急救医学是一门多专业的综合学科，是处理和研究各类急性病发病阶段和因此可能发生的急性器官功能衰竭、慢性病急性变、各种急性创伤的初步处理，以及抢救垂危重症病人生命的一门专业学科。它不处理疾病的全过程，而把重点放在病伤的急性阶段，为以后的诊治创造良好的条件。由于急救医学的内容不仅仅局限在医院内，甚至更多的是在医院外发生，因此急救医学的内容也包括有关通信、现场抢救、运输等方面的问题。

急救医学需要科学研究，它是急救医学在科学原则的基础上进一步发展所必需的。为此需要在多水平上开展急救医学科学研究，包括从细胞到人群、从动物试验到临床试验等各类科学研究。医学科学研究有多种分类形式，常见的是从研究形式上将医学研究分为观察性研究(Observational Study)和试验性研究(Experimental Study)。

观察性研究又称为调查研究，是指研究者不对研究因素进行人为控制，而是通过现场调查分析的方法进行的研究，可分为横断面研究、病例-对照研究和队列研究。横断面研究又称为现况研究，主要用于描述疾病或健康状态在人群、时间、地区的分布，并可依据调查样本数据估计总体参数，也可以为病例-对照和队列研究提供病因线索；病例-对照和队列研究又称为分析性研究，主要用于证实疾病的各种危险因素，估计危险因素对疾病作用的大小，并提出可能的干预措施，从而探索和验证病因假说。

观察性研究的主要特点是研究对象及其相关因素(包括研究因素和非研究因素)是客观存在的，并且不能通过随机化分组来均衡混杂因素对研究结果的影响，研究者只能被动地对研究对象进行观察，并客观地记录观察结果。因此，观察性研究结果的分析和解释要非常谨慎。

试验性研究是指研究者根据研究目的人为地设置干预措施，按照重复、对照、随机化的基本原则控制非干预因素对试验结果的影响，通过对试验结果的分析，评价干预措施的效果。试验性研究的优点在于能够较好地控制非试验因素的影响，避免人为造成的偏倚，使比较组间具有较好的可比性。

从研究性质来划分，临床试验属于试验研究的范畴。临床试验的主要特点是以人为研究对象，其目的是证实或揭示治疗方法或预防措施的疗效和安全性，综合评价治疗方法或预防措施的效果

和价值。临床试验作为一种以“人”为受试对象的试验研究,必须面对医学伦理学的问题,这一点研究者尤其值得注意。

本章的分析案例取材于近年来发表在《中国危重病急救医学》杂志的科研论文,案例的选择原则是所选论文的类别在该杂志发表的文章中占有较高比例,是危重病急救医学领域常见的研究形式。案例分析时,首先按照研究类型对文章进行分类,即临床试验、观察性研究和试验性研究,针对每种类型的研究主要从研究设计、统计分析方法应用及统计结果的解释和表达等方面进行剖析,使读者能够结合具体问题,深刻体会危重病急救医学研究中的统计学思想和统计方法的正确应用。为了便于读者完整地理解医学研究中统计设计和数据处理时需要特别注意的问题,本章在结合案例进行分析时,不仅分析了文章中存在的统计学问题与不足,同时也对统计方法运用得当之处给予肯定。

2.2 临床试验研究中常见的统计学问题及案例辨析

临床试验是危重病急救医学的主要研究类型之一,《中国危重病急救医学》杂志发表的许多论文属于此类研究。本节内容主要通过案例分析,将临床试验相关的统计学理论贯穿于案例辨析之中,使读者能够更好地理解临床试验中的统计设计与数据分析方法,从而在临床科研实践中正确运用统计学理论与方法。本章所选的分析案例取材于近几年发表在《中国危重病急救医学》杂志的文章,该类文章在杂志中占有较大的比例,具有较好的代表性。

【案例2-1】 目的是探讨低分子肝素依诺肝素用于普通外科手术后预防深静脉血栓形成(DVT)和肺栓塞(PE)的有效性和安全性。以随机双盲方法,将1928例外科手术患者根据术后抗凝与否分为观察组和对照组。观察组960例,给予依诺肝素干预,对照组968例不予干预。观察两组患者DVT、PE以及出血事件的发生情况。

【案例辨析与释疑】 一项医学研究的结果是否具有科学性,取决于研究的设计是否合理、数据分析方法是否恰当、结果表达是否规范以及结果解释是否合理,下面对所选案例从上述几个方面进行辨析。

关于临床试验设计:解读论文“低分子肝素预防外科术后患者的肺栓塞”的摘要可以看出,该研究的设计类型属于“随机对照双盲临床试验设计”,此类研究的研究结果是最理想的临床证据。在条件允许的情况下,研究者应尽可能选择此类研究,因为它不仅能够通过随机化分组均衡非干预因素对试验组和对照组的影响,而且通过“双盲法”有效地控制了研究者和受试者的主观因素和心理因素对试验结果的影响,从而使试验结果更加可靠,组间具有良好的可比性。

从试验设计的三个基本要素来看,本研究的“受试对象”为“外科手术患者”,“试验因素”为“皮下注射依诺肝素与否”,“试验效应”为研究者根据专业知识确定的能够反映干预效果的一系列指标。该试验设计遵从“对照、重复、随机化”的试验设计前三个基本原则,即设立对照组、随机化分组、适宜的样本含量,从而使对比组之间满足很好的均衡性。值得肯定的是该研究还采用了临床试验研究中倡导的“双盲法”,并且注意到开展临床试验的伦理学要求,值得临床试验研究者借鉴。这是因为临床试验的受试对象是人,因此必须面对医学伦理学的问题。要求研究者在临床试验开始之前,应该将临床试验的目的、方法、预期结果和可能出现的副作用如实地告诉每名受试者,使其在充分理解并进行考虑之后自由选择是否参与临床研究,也就是“知情同意”,这充分体现了对受试者尊严和自主权的尊重。“失访”也是临床试验研究者需要认真考虑的问题。若患者失访较多,会直接影响试验结果的可信度。因此,在临床试验设计时,研究者必须制定有效措施提高患者的依从性,降低失访率。从《中国危重病急救医学》杂志上发表的临床试验类文章来看,已经有一

部分作者在文章中有关于“医学伦理学”方面的说明,但还有许多作者并未提及。多数文章未对“患者依从性”进行说明。这一点需要引起临床研究者的注意。

关于统计分析方法应用:统计分析方法的应用主要体现在统计描述指标的选择和假设检验方法的选择上。统计描述指标的选择取决于资料的类型及其分布。如若测量指标是定量资料,且分布为正态分布或近似正态分布,则应该选择“算术均数”描述其集中趋势,“标准差”描述其离散程度,即用 $\bar{X} \pm S$ 来描述;若计量资料呈偏态分布或分布类型未知,则应该选择“中位数”描述其集中趋势,“四分位数间距”描述其离散程度。若资料为定性资料,则应根据实际情况选择“率”或“构成比”对资料进行统计描述。

本例中,研究者所选指标可分为两组,一组为基线资料指标,包括患者的性别、年龄、体质指数、手术时间、肥胖、静脉血栓病史、静脉曲张、慢性阻塞性肺疾病史、慢性心功能不全病史、激素治疗史、手术类型等;另一组为疗效指标,包括 DVT 发生率、PE 发生率、出血事件发生率等。基线资料中年龄、体质指数为定量资料,其余均为定性资料;疗效资料均为二值或多值有序资料。

文章中统计学方法描述为:“应用 SPSS 11.0 软件进行数据分析;计量资料以中位数(M)表示,组间比较采用两独立样本 t 检验;定性计数资料采用 Fisher 确切概率法或 χ^2 检验; $P < 0.05$ 为差异有统计学意义。”

可以看出,上述定量资料的统计学处理存在明显的错误。文章中对年龄、体质指数和手术时间采用“中位数”描述其集中趋势,说明这三个指标的资料呈非正态分布,不满足 t 检验的条件,因此,定量资料组间比较的假设检验方法选择是错误的,此种情况下,应选择非参数检验方法——Wilcoxon 秩和检验。若用“中位数”描述平均水平,还应给出相应的离散指标“四分位数间距”,因此,定量资料的统计描述也是不全面的。

定性资料的描述指标是率或构成比,假设检验方法的选择应符合相关要求。四格表资料假设检验方法的选择条件为:

当 $n \geq 40$ 且所有理论频数均大于 5,用 χ^2 检验的基本公式计算检验统计量;

当 $n \geq 40$ 且有 $1 \leq T < 5$ 时,用 χ^2 检验的校正公式计算检验统计量;

当 $n < 40$ 或 $T < 1$ 时,用四格表资料的 Fisher 确切概率法。

其中 n 为样本例数, T 为理论频数。

对于 $R \times C$ 列联表资料:

当 $T > 1$ 且 $1 \leq T < 5$ 的格子数不超过格子总数的 $1/5$ 时,用 χ^2 检验的基本公式;

当 $T < 1$ 或 $1 \leq T < 5$ 的格子数超过格子总数的 $1/5$ 时,用 $R \times C$ 表资料的 Fisher 确切概率法(可用 SAS 软件实现)。

本例定性资料的假设检验方法应根据上述条件选择适宜的方法。SPSS 的输出结果会有相应的提示,研究者可根据提示选择适宜的输出结果。当然,对于结果变量为多值有序变量的疗效资料,可采用秩和检验来比较不同组间疗效有无差异。

关于统计结果的表达:统计结果表达的有效工具是统计图表。统计表用简明的表格形式,有条理地罗列数据和统计量,方便阅读、比较和计算,可以代替冗长的文字叙述。统计图能够通过图形使统计数据表达形象化、直观化,易于作分析比较,并且能够给读者留下深刻的印象。原作者对研究结果用统计表来表达,形式摘录如下。

由表 2-1 可以看出,患者一般资料比较的指标较多,既有定量资料,也有二分类和多分类的定性资料。所有指标在一张统计表中表达显得内容过多,也会给统计表的标目设计和线条应用带来

不便。如本例中横标目“组别”在上下两段表格中重复出现,线条也不符合统计表的编制要求。总标目“性别”应改为“性别[例(%)]”,并且应给出相对数指标。另外,“年龄”、“体质指数”、“手术时间”的描述还应该增加“四分位数间距”。改进方法:可以按照指标的性质或表达形式将指标分为2~3组,上述表格可据此拆分为3~4张表格,这样即可使表格编制变得简单、规范,也使结果具有更好的对比性和可读性。修改后的表格结构见表2-2~表2-5。

表2-1 是否用依诺肝素治疗两组患者的一般资料

组别	例数	例数		年龄 (M , 岁)	体质指数 (M , kg/m^2)	危险因素[例(%)]					
		男	女			肥胖	静脉血栓病史	静脉曲张	COPD病史	慢性心功能不全病史	激素治疗史
对照组	968	535	433	63.8	21.0	78(8.0)	26(2.7)	35(3.6)	22(2.3)	28(2.9)	12(1.2)
观察组	960	546	44	64.0	20.8	80(8.3)	28(2.9)	32(3.3)	20(2.1)	25(2.6)	9(0.9)

组别	例数	手术类型[例(%)]							非恶性肿瘤手术 [例(%)]	恶性肿瘤手术 [例(%)]	手术时间 (M , h)
		胃小肠	肝胆胆道	结直肠	静脉曲张	脾切除	胰腺炎	其他			
对照组	968	405 (41.8)	186 (19.2)	216 (22.3)	52(5.4)	36(3.7)	32(3.3)	41(4.2)	353(36.5)	615(63.5)	2.3
观察组	960	410 (42.7)	177 (18.4)	208 (21.7)	48(5.0)	38(4.0)	30(3.1)	49(5.1)	342(35.6)	618(64.4)	2.2

表2-2 两组患者一般资料比较

组别	例数	例(%)		年龄(岁)		体质指数(kg/m^2)	
		性别: 男	女	M	$Q_1 \sim Q_3$	M	$Q_1 \sim Q_3$
对照组	968	535(·)	433(·)	63.8	·	21.0	·
观察组	960	546(·)	44(·)	64.0	·	20.8	·

表2-3 两组患者危险因素比较[例(%)]

组别	例数	肥胖	静脉血栓病史	静脉曲张	COPD病史	慢性心功能不全病史	激素治疗史
对照组	968	78(8.0)	26(2.7)	35(3.6)	22(2.3)	28(2.9)	12(1.2)
观察组	960	80(8.3)	28(2.9)	32(3.3)	20(2.1)	25(2.6)	9(0.9)

表2-4 两组患者手术类型比较[例(%)]

组别	例数	胃小肠	肝胆胆道	结直肠	静脉曲张	脾切除	胰腺炎	其他
对照组	968	405(41.8)	186(19.2)	216(22.3)	52(5.4)	36(3.7)	32(3.3)	41(4.2)
观察组	960	410(42.7)	177(18.4)	208(21.7)	48(5.0)	38(4.0)	30(3.1)	49(5.1)

表2-5 两组患者肿瘤类型及手术时间比较

组别	例数	例(%)		手术时间(h)	
		肿瘤类型: 恶性	良性	M	$Q_1 \sim Q_3$
对照组	968	353(36.5)	615(63.5)	2.3	·
观察组	960	342(35.6)	618(64.4)	2.2	·

关于统计结果的解释:统计结果的解释主要集中在“讨论”部分,原作者往往要在“讨论”部分引用统计结果作为支持其新发现、新结果和新观点的统计学依据。在解释统计结果时,原作者应注意区分“统计显著”和“临床疗效显著”两种不同的结论,不能将两者混为一谈。当 $P < 0.05$ 时,我们下的结论是“组间比较差异有统计学意义”, P 值越小就越有理由拒绝“组间无差异”的假设,但它并不能说明组间临床实际差异的大小。理论上讲,当样本含量足够大时,很容易得到“差异有统

计学意义的结论”，但差异是否在临床上有意义还需要结合专业知识来判断。如评价某降压药物的疗效，统计结论为“试验组和对照组疗效差别有统计学意义($P < 0.05$)”，但试验结果显示，两组均数差仅为 2mmHg。若临床专业认为，降压药的降压效果平均在 5 mmHg 以上才有临床意义，那么，此时的“差别有统计学意义”就不能认为是“临床疗效显著”。可见，有统计学意义的检验结果并不意味着确有疗效。

本例中原作者将研究结果解释为“本研究中证实了在住院期间，观察组 DVT 发生率较对照组明显下降(2.9% 比 6.1%)，PE 发生率显著低于对照组(0.3% 比 1.4%)。”可以看出，原作者对“差异有统计学意义”的解释用到了“明显下降”、“显著低于”之类表述方式。此种表达只有在两组样本率的差别程度已经达到了临床上认可的“明显”或“显著”的程度才能使用，否则应尽量避免使用此种表达方式。

【案例 2-2】 目的是观察中西医结合治疗老年重症肺炎的疗效。采用随机化分组方法，将受试对象分为试验组(51 例)和对照组(49 例)。对照组采用常规疗法，试验组在常规疗法的基础上加用痰热清注射液 20 ml 和参附注射液 80 ml 静脉滴注。观察指标有总有效率、病死率、住院天数、体温、呼吸频率(RR)、心率(HR)、白细胞计数(WBC)、动脉血氧饱和度(SaO_2)、氧合指数($\text{PaO}_2/\text{FiO}_2$)等。统计学方法：用 SPSS 12.0 统计软件，计量资料以均数 \pm 标准差($\bar{x} \pm s$)表示，采用 t 检验，计数资料用 χ^2 检验， $P < 0.05$ 为差异有统计学意义。

【案例辨析与释疑】 从上述摘要信息可知，本研究属于“随机对照临床试验”，受试对象为满足纳入条件的老年重症肺炎患者，分为两组，对照组采取常规疗法，试验组采取常规疗法 + 中药制剂，两组同期平行观察，组间差别反映了中西医结合疗法与常规疗法的疗效差异。由于是临床试验，原作者还应该考虑相关的伦理学问题，如患者是否知情同意，是否有医学伦理委员会的许可等。

统计描述：从观测指标来看，属于定量资料的指标有住院天数、体温、呼吸频率(RR)、心率(HR)、白细胞计数(WBC)、动脉血氧饱和度(SaO_2)、氧合指数($\text{PaO}_2/\text{FiO}_2$)；属于定性资料的指标有总有效率和病死率。原作者对定量资料统计描述指标的选择应考虑资料的分布类型，而不能全都采用 $\bar{x} \pm s$ ，如一般情况下“住院天数”多为偏态分布资料，“氧合指数”一般也非正态分布。若指标原始数据的分布为非正态，就需要用中位数和四分位数间距来描述。

假设检验：假设检验方法的选择取决于资料的分布类型。本例总有效率和病死率的比较采用 χ^2 检验是适宜的，但定量资料的假设检验方法欠妥。本例定量资料的统计结果摘录见表 2-6。

表 2-6 两组老年重症肺炎患者治疗前后检测指标变化比较($\bar{x} \pm s$)

组 别		例 数	体温 ($^{\circ}\text{C}$)	RR (次/min)	HR (次/min)	WBC ($\times 10^9/\text{L}$)	SaO_2	Pa/FiO_2 (mmHg)
对照组	治疗前	51	38.10 ± 2.09	29.81 ± 4.52	112.17 ± 18.09	15.48 ± 7.39	0.766 ± 0.105	275.86 ± 70.57
	治疗后	46	36.5 ± 20.76	21.17 ± 4.21	85.64 ± 17.25	8.05 ± 2.14	0.946 ± 0.046	333.49 ± 65.60
观察组	治疗前	49	38.20 ± 2.13	29.57 ± 4.82	113.33 ± 19.08	15.38 ± 7.45	0.756 ± 0.107	287.47 ± 67.89
	治疗后	39	37.63 ± 0.72	23.73 ± 5.66	89.03 ± 18.55	9.37 ± 2.80	0.952 ± 0.037	323.30 ± 59.34

从上述记录结果可以看出，该资料属于前后测量的重复测量设计资料。本例中，原作者采用 t 检验的方法对组内前后测量及组间相应时间点的测量结果进行多次重复 t 检验，必然会增大犯假阳性错误的概率，而且也不能分析处理因素与时间因素之间的交互效应，因此是不适宜的。此类资料可采用重复测量设计定量资料的方差分析进行组间差别的比较。

重复测量设计资料是医学研究中常见的资料类型，现以表 2-7 数据为例，说明重复测量数据的特征及常用的统计分析方法。

表 2-7 12 名患者手术前后症状评分

患者编号	处理分组	症 状 评 分				
		手术前	*：10 天	2 个月	4 个月	6 个月
1	A	0.60	0.67	2.84	2.10	2.00
2	A	1.42	3.40	4.10	2.92	2.65
3	A	0.90	2.30	2.70	1.70	1.10
4	A	1.10	1.40	1.00	2.60	0.90
5	A	2.30	2.20	3.80	3.50	2.50
6	A	0.81	1.20	1.12	1.61	1.49
7	B	1.20	1.10	1.13	3.49	1.57
8	B	2.71	2.04	2.61	2.17	2.15
9	B	1.02	1.43	1.61	1.70	2.82
10	B	0.85	1.25	1.66	2.13	1.04
11	B	0.60	2.50	2.20	1.20	1.11
12	B	0.90	0.80	0.70	1.00	0.80

注：* 代表“手术后时间”。

表 2-7 的数据记录了接受不同手术方式(处理因素,两个水平:A 处理,B 处理)的 12 例患者分别在术前及术后不同时间点的症状评分,这类数据具有以下特征。

(1) 每个受试对象的指标在不同的时间点上被重复测量,得到的数据在形式上类似于“随机区组设计的数据”,即每个受试对象在不同时间点的测量结果形成一个区组,但两者又有明显的区别:随机区组设计每个区组内的试验单位是彼此独立的,不同处理只能在区组内随机分配,而重复测量设计中的“处理”是在区组间(受试对象)进行的,区组内的各时间点是固定不变的,不能随机分配。如表 2-7 中 A、B 两种不同的手术方式随机分配给每位患者后,每位患者症状评分测量的时间点是固定的。

(2) 与随机区组设计相比,重复测量设计资料的另一个特征是区组内测量结果(受试者不同时间点的测量值)彼此不独立,也就是说同一受试者在不同时间点上的测量结果是高度相关的。

如果重复测量资料按照随机区组设计资料的分析方法去分析显然是不适宜的,除非资料满足“球对称”假设。一般统计分析软件如 SPSS 在做重复测量资料分析时,同时会给出“球对称”假设检验的结果。有些研究者为了避免重复测量数据特殊性,只分析最后一次的测量结果,这样就会丧失很多的“过程”信息,如通过记录不同时间点的重复测量数据,可以了解测量指标的时间变化趋势、药物在血液中浓度变化的时间分布特征、不同时间的治疗效果等。因此,对于重复测量设计定量数据,应该用重复测量资料的单变量方差分析(ANOVA)或多变量方差分析(MANOVA)方法来处理,其中 ANOVA 是一种比较容易理解、统计分析结果比较明确的一种方法。关于重复测量设计定量资料方差分析的基本思想在此不再赘述,有兴趣的读者可参阅孙振球主编、供研究生使用的《医学统计学》教材。近年来,采用混合效应方差分析模型处理重复测量设计定量资料更为多见。

【案例 2-3】 目的是探讨应用“特殊部位脑出血微创定位器”和微创术救治大量脑干出血患者的疗效。受试者为 8 例严重脑干出血患者,干预措施为“特殊部位脑出血微创定位器”引导下的微创术,疗效测量呼吸、血压、心率等生命体征指标,以及影像检查结果。

【案例辨析与释疑】 本研究属于临床试验,目的在于探讨应用自制的“特殊部位脑出血微创定位器”和微创术(干预措施)救治大量脑干出血患者的疗效。但从试验设计来看,本例只有治疗组,未设立对照组,导致干预效果的判断缺乏参照,无法做出是否有效或疗效优劣的判断。从样本

量来看,本例仅对8例严重脑出血患者实施了干预措施,样本量较小,观察结果统计描述的意义不大。本例中,结果描述多采用个案陈述的方式,未做统计学处理,无法对观察结果进行推论,因此在结果解释方面要非常谨慎。

原作者在讨论部分对观察结果解释为:“采用自制的特殊部位脑出血简易立体定向仪,对出血量在10 ml以上的8例脑干出血患者进行微创抽吸术,结果6例患者术后生命体征有不同程度改善,其中3例在抽吸40%~50%的血肿后自主呼吸逐渐提高,2例术后72 h清醒,2例术后7 d清醒。提示大量脑干出血采用微创术抽吸血肿疗效显著,不仅能显著减低病死率,而且有望明显降低致残程度。”

很显然,由个案观察结果陈述得出干预疗效显著、显著降低病死率等结论是缺乏科学数据支撑、无法令人信服的。

在《中国危重病急救医学》杂志中,有许多文章属于此类单组、无对照的小样本临床观察。此类观察结果作为临床证据的可信度远远低于随机对照临床试验。若要得到干预措施效果评价的科学证据,还需要通过随机对照临床试验获得。

2.3 观察性研究中常见的统计学问题及案例辨析

由于受到医学伦理、传统观念及研究经费等方面的限制,许多临床医学研究很难达到试验研究的条件,只能通过观察性研究来观察研究因素与效应之间是否存在关联以及关联程度的大小。观察性研究中由于对研究因素不进行人为控制,因此具有易于操作、便于实施的特点,而且多数是在自然条件下进行的,对于目标人群来说研究对象具有较好的代表性。另一方面,也正是由于对研究因素不进行人为控制,使得观察性研究的观察结果无法排除混杂因素带来的偏倚,研究结果的可重复性降低。

《中国危重病急救医学》杂志中发表的许多文章属于临床观察性研究。本节将选取部分有代表性的临床观察性研究案例,从研究设计、数据分析及结果解释等方面,对不同类型观察性研究中常见的问题进行辨析,以便读者更好地掌握危重病急救医学中临床观察研究设计和数据分析的关键环节,避免再犯类似的错误。

【案例2-4】 目的是了解某医院院内感染发生情况、抗菌药物使用情况及病原体检出情况。调查对象为某医院2010年8月24日00:00~24:00所有住院患者,主要统计指标有医院感染现患率、感染部位构成比、感染病原菌构成比、抗菌药物使用率等,研究结果为本次调查时点的样本统计量。

【案例辨析与释疑】 本案例属于观察性研究中的横断面调查,也称为现况调查,是常用的描述性流行病学方法之一。它是按照事先设计的要求在某一人群中应用普查或抽样调查的方法,收集特定时间疾病的描述性资料,以描述疾病的分布及观察某些因素与疾病之间的关联。从时间上来说,横断面调查是在某一时点或在短暂的时间内完成的。因此,对于横断面调查,研究者首先应该明确调查的目标总体是什么?调查的方法是普查还是抽样调查?若为普查,则调查结果只需给出描述指标即可,若是抽样调查,不仅要给出样本统计量,还应该给出总体参数的置信区间,这样,统计描述才是全面的,也便于同类调查结果之间的比较和解释。

本例中调查结果解释为:本次调查的医院感染现患率为4.7%,例次现患率为5.4%,与任南等报道的结果相似,但高于本院以往日常连续性监测结果(2009年全年医院感染率为2.2%,例次感染率为2.5%;2010年1月至8月平均医院感染率为2.1%,例次感染率为2.5%)。

本例中,研究者调查了特定时点某医院所有住院患者的医院感染和抗生素的使用情况。由于

住院患者具有较强的流动性,而且季节因素对院内感染有明显地影响,因此特定时点的所有住院患者可以看作是该院住院患者的一个样本,调查结果是样本统计量。若直接用本次调查的样本统计量估计该院的年度院内感染现患率欠妥当。建议在统计结果描述中,同时给出相应的置信区间。若直接用本次调查的样本统计量与其他调查结果比较,则不能得出总体有差别的结论。

【案例2-5】 目的是调查广州市院前急救患者流行病学情况,探讨院前急救病例特点。数据来源于广州市急救医疗指挥中心信息统计数据库中2008年度全部数据。结果描述了急救呼叫的日时间分布;急救车出车次数的地区分布;院前急救患者的性别、年龄、疾病、死因等分布情况。结论:加强心脑血管疾病和呼吸系统疾病防治,提高中老年患者的常见急危重症早期识别与院前急救水平对降低死亡有重要意义;加强安全生产、遵守交通法规、强化法制意识是降低创伤,尤其是交通意外发生及其死亡的有力手段。

【案例辨析与释疑】 本例中,原作者对来源于某急救中心一定时期内的院前急救患者的临床资料进行了统计描述与分析,如对患者性别、年龄等基本资料的描述,对可能影响急救效果的影响因素(急救呼叫日时间分布、急救车出车次数的地区分布、疾病诊断)的描述,以及对救治效果的描述(死亡原因)等,属于“回顾性观察研究”,其目的在于通过院前急救患者临床资料的汇总描述,了解疾病相关因素的分布情况,尤其是可能影响救治效果的因素。此类文章在《中国危重病急救医学》杂志中很多,如“肺血栓栓塞症252例临床诊治体会”、“12例心力衰竭致大量胸腔积液患者临床分析”、“50例急性肺栓塞临床诊治分析”、“机械通气下内镜治疗食管胃底静脉曲张破裂出血的救治体会”、“外膜肺氧合治疗急性呼吸窘迫综合征的临床疗效观察”等。这些文章一般是临床工作者对一定时期内特定病例临床资料的总结性描述,研究结果一般也仅限于了解患者的疾病特点或诊疗救治规律,初步观察某些因素与疾病或健康状况之间的关系,从而为建立病因假说、评价诊疗措施的效果提供一些参考信息。但是这类研究结果若作为临床证据用于指导疾病诊疗,其科学性和可用性需要认真考虑。

首先是样本的代表性问题。这类文章的数据一般是“某一所医院”或医疗机构“一定时期”的病例资料,尤其是在样本量较小的情况下,样本的代表性是比较差的;其次,在数据分析时,虽然研究者对患者的人口学特征(年龄、性别等)、疾病特征(症状体征、检查检验结果等)、暴露因素或可能的危险因素的分布情况进行了描述,但由于缺乏对照,读者很难就此数据明确判断某因素是否与疾病的发生、进展或转归有关。

本例的观察样本来自某市的急救中心,且样本量较大(97823例),样本具有一定的代表性,通过描述性分析,在一定程度上能够反映该市院前急救患者的人群特征、疾病特征及救治转归情况,如院前急救患者疾病类别分布、年龄分布、院前急救死亡患者的死因分布等,这些关键因素的分布特点能够为院前急救的重点人群、主要疾病及主要死因等的识别提供有价值的线索。但是,由于观察性研究的混杂因素较多,若要进一步了解哪些因素与疾病发生或转归有关,还需要对临床观察研究提供的病因线索,进行“病例-对照研究”。

“病例-对照研究”是分析流行病学最基本、最重要的研究方法,其基本原理是以现在确诊的患有某种疾病的患者作为病例,以不患该疾病但具有可比性的个体作为对照,通过询问、试验室检查或复查病史,收集既往各种可能的危险因素的暴露史,测量并比较病例组与对照组的暴露比例,经统计学检验,若两组差别有统计学意义,则可认为因素与疾病之间存在统计学上的关联。“病例-对照研究”中表示疾病与暴露之间关联强度的指标是比值比,即OR(Odds Ratio)值,也称优势比或比数比。但在《中国危重病急救医学》杂志中,却很少见到采用“病例-对照”的研究方法来分析影响疾病发生或转归的因素。有关“病例-对照研究”的设计及统计分析,读者可参阅流行病学相关书籍,此处不再赘述。

【案例 2-6】 目的是评价血清性激素水平在腹部外科疾病伴全身炎症反应综合征(SIRS)患者病情预后评估中的应用价值。采用前瞻性、单中心、观察研究方法。选择 2009 年 7 月 10 日至 2010 年 2 月 9 日天津市南开医院外科重症监护病房(SICU)收治的 39 例腹部外科疾病伴 SIRS 患者,根据入院 72 h 内的血清雌二醇、泌乳素及睾酮水平,将患者分为雌二醇升高组和未升高组、泌乳素升高组和未升高组、睾酮降低组和未降低组,随访 28 天,分别观察各组患者急性生理学与慢性健康状况评分系统Ⅱ(APACHEⅡ)评分、总住院时间、住 ICU 时间、总住院费用及 28 天转归。

【案例辨析与释疑】 本例中,研究者将某医院一定时间段内符合入选条件的病例纳入研究队列,并按照病例进入研究队列时的激素水平将其分为暴露组(高雌激素水平组、高泌乳素水平组、高睾酮水平组)和非暴露组(低雌激素水平组、低泌乳素水平组、低睾酮水平组),随访观察 28 天后,观察病例的转归(28 天病死率),以此来分析暴露因素(高激素水平)与疾病转归之间的关系,属于观察性研究中的“队列研究”。

“队列研究”也称为前瞻性研究、随访研究或纵向研究,其基本原理是选定暴露于及未暴露于某因素的两组人群,随访观察一定的时间,比较两组人群某种疾病的结局(一般指发病率或死亡率),从而判断该因素与疾病结局有无关联及关联大小的一种观察性研究方法。队列研究可分为三种类型。

(1)前瞻性队列研究:研究对象的确定与分组是根据队列研究开始时的状态确定的,研究的结局需要随访观察一段时间才能得到。其优点是研究者可以直接获取第一手资料,资料的偏性较小。但若观察人群过大、观察时间过长,则会影响其可行性。

(2)历史性队列研究:研究工作是现在开始的,研究对象是过去某个时间进入队列的,研究对象的确定与分组是根据研究对象进入队列时的暴露情况进行的,研究的结局在研究开始时已经发生,暴露到结局的方向是前瞻性的,但研究工作的性质是回顾性的。历史性队列研究的优点是省时、省力,出结果快,因而适用于长诱导期或长潜伏期疾病的病因分析。但这种设计类型常常缺乏影响暴露于疾病关系的混杂因素的资料,以至于影响暴露组与非暴露组的可比性。

(3)双向队列研究:是指历史性队列研究之后,继续进行前瞻队列研究。这种研究具有上述两种研究的优点,在一定程度上弥补了它们的不足。

队列研究是根据研究对象是否暴露于某影响因素分组的,而暴露因素是客观存在的(如患者激素水平的高低),并非研究者人为设置的干预因素,因此,队列研究不能进行随机化分组,这也是观察性研究和试验性研究的主要区别之一。由于不能实现随机化分组,队列研究的组间均衡性检验就显得十分重要。若组间均衡性差,队列研究的观察结果的组间差异则无法用暴露因素来解释,这一点在数据分析及结果解释中应特别注意。

统计分析与结果解释:本文的统计学处理描述为:“采用 SPSS 16.0 统计软件进行数据分析。符合正态分布的定量资料以均数 \pm 标准差($\bar{x} \pm s$)表示,采用独立样本 t 检验;定性资料采用 χ^2 检验, $P < 0.05$ 表示差异具有统计学意义”。部分观察结果描述见表 2-8。

表 2-8 雌二醇升高组与未升高组经腹部外科疾病伴全身炎症反应综合征患者临床数据比较

组别	例数	年龄 ($\bar{x} \pm s$, 岁)	APACHE II 评分 ($\bar{x} \pm s$)	总住院时间 ($\bar{x} \pm s$, d)	住 ICU 时间 ($\bar{x} \pm s$, d)	总住院费用 ($\bar{x} \pm s$, 万元)	28d 病死率 (%)
升高组	20	47.3 \pm 23.5	11.6 \pm 7.8	24.3 \pm 11.6	8.5 \pm 4.0	10.5 \pm 0.83	10.0
未升高组	19	50.2 \pm 22.5	15.2 \pm 8.8	22.3 \pm 13.1	12.1 \pm 7.9	6.20 \pm 0.35	35.7 ^a

从上面的数据分析方法和结果表达形式可以看出:本文仅从单因素分析的角度,对暴露组和非暴露组的结局指标进行了描述和假设检验。其中混杂因素只分析了一个变量“年龄”,但可能影响疾病结局的重要因素还应包括“性别、两组的疾病分布、住院期间的治疗方法等”,这些变量均会对

疾病结局产生明显的影响,但文中并未对这些因素进行组间均衡性检验,因此,即使假设检验结果有统计学意义(如28天病死率),也不能就此说明这种差异是由于激素水平不同引起的。而本文却基于上述结果做出了“不同性别患者中雌激素水平的升高或降低与患者预后关系密切,在腹部外科手术伴SIRS患者中,早期雌激素水平升高者明显好于未升高者”的结论,显然是不适宜的。

2.4 试验研究中常见的统计学问题及案例辨析

试验性研究是指研究者根据研究目的人为地设置干预措施,按照重复、对照、随机化的基本原则控制非干预因素对试验结果的影响,通过对试验结果的分析,评价干预措施的效果。与观察性研究相比,试验性研究的优点在于能够通过随机分组,较好地控制非试验因素的影响,使组间具有较好的可比性。广义来讲,前文所提及的“临床试验”也属于试验研究。但本节所述的试验研究主要是指以动物、标本或其他生物材料为研究对象的试验性研究,其中动物试验是最常见的试验研究类型。文中的分析案例取材于《中国危重病急救医学》杂志近年来发表的试验研究论文。本文将结合所选案例,分析试验研究的统计设计和数据分析中的常见问题及正确处理方法,尽量避免空洞的理论讲解,以期达到学以致用目的。

【案例2-7】 目的是研究N-乙酰半胱氨酸(NAC)对大鼠肺缺血/再灌注损伤(LIRI)中细胞凋亡及天冬氨酸特异性半胱氨酸蛋白酶3(caspase-3)表达的影响,并探讨其可能机制。方法:雄性SD大鼠42只,按随机数字表法均分为假手术组、LIRI组、NAC组3组,各组分别于3h、6h取肺组织,通过流式细胞仪检测凋亡细胞并计算凋亡率,测定丙二醛(MDA)浓度和超氧化物歧化酶(SOD)活性(黄嘌呤氧化酶法),检测caspase-3 mRNA表达,电镜下观察肺脏超微结构的改变。

【案例辨析与释疑】 本研究属于动物试验研究,试验设计的三要素分别是:受试对象(雄性SD大鼠)、处理因素(两个因素为NAC和时间,NAC有三个水平:假手术、LIRI模型、NAC+LIRI模型;时间有两个水平:3h、6h)、试验效应(细胞凋亡率、MDA浓度、SOD活性、caspase-3 mRNA表达、肺脏超微结构);试验设计遵循试验设计的三个主要原则,即对照(设立对照组)、重复(一定的样本含量)、随机化(随机化分组),因此,从统计学角度来说,本研究设计符合试验设计的基本要求。

与临床试验相比,动物试验的最大优点就是许多因素可以由试验者自行安排,如制作各种动物模型,对动物种属、体重的限制等。因此,动物试验的设计可以更加严密,使对比组之间的可比性更强,这样就可以最大程度的控制非试验因素对试验结果的影响。但需要特别注意的是,动物和人对同一处理的反应既有共同性也有差异性,因此,动物试验的结果不能直接用来解释人类的现象。本例的主要试验结果表达见表2-9。

表2-9 NAC对大鼠LIRI后各时间点肺组织凋亡率、MDA、SOD、caspase-3 mRNA表达的影响 $\bar{x} \pm s$

组别	例数	凋亡率(%)		MDA(nmol/mg)		SOD(U/mg)		caspase-3 mRNA	
		3h	6h	3h	6h	3h	6h	3h	6h
假手术组	14	2.19±0.48	2.55±0.36	0.73±0.23	1.08±0.42	60.51±6.81	64.19±6.60	0.216±0.046	0.227±0.037
LIRI组	14	25.60±3.22	26.01±4.50	3.26±0.32	3.53±0.46	32.80±3.82	33.44±3.24	0.717±0.037	0.744±0.046
NAC组	14	14.42±1.61	10.02±1.64	1.75±0.33	2.15±0.25	42.76±2.06	44.94±3.11	0.441±0.038	0.410±0.037

从表2-9可以看出,测量的4个指标均为定量资料,原作者采用 $\bar{x} \pm s$ 来进行统计描述,但需要强调的是,若用 $\bar{x} \pm s$ 来描述资料的分布特征,前提条件是该资料应服从正态分布,本例中的“细胞凋亡率”符合正态分布吗?若资料为偏态分布或分布类型未知,则宜用中位数和四分位数间距来描述其分布特征。

对于三个处理组间的比较，原作者描述为“各组间应用方差分析检验方差齐性，方差齐者进行组间两两比较采用 t 检验。 $P < 0.05$ 为差异有统计学意义”。

结合表 2-9 的结果，我们来分析本文中的统计学处理存在的问题。首先，该试验有三个分组，每组的四个测量指标分别在两个不同时间点进行，得到的资料属于重复测量设计资料。本例原作者采用 t 检验的方法对组间相应时间点的测量结果进行多次重复 t 检验，必然会增大犯假阳性错误的概率，因此是不适宜的。此类资料可采用重复测量设计资料的方差分析进行组间差别的比较。有关重复测量设计资料的资料特征及统计分析方法的内容，参见【案例 2-2】，此处不再重复。

【案例 2-8】 目的：观察阿司匹林对原代培养大鼠 II 型肺泡上皮细胞(AEC II)的保护效应，并探讨其抗氧化损伤的机制。方法：将原代分离、纯化、培养的离体大鼠 AEC II 分为 5 组。过氧化氢损伤组、生理盐水(NS)组、阿司匹林预处理组，3 h 后观察细胞形态学变化和贴壁细胞计数，检测指标为细胞存活率、血红素氧合酶-1 的蛋白及 mRNA 表达。

【案例辨析与释疑】 本试验的研究对象为离体大鼠II型肺泡上皮细胞(AECII)，属于生物材料类。处理因素共有 5 个水平，形成 5 个试验组，处理效应通过测定相关指标表达，见表 2-10 和表 2-11。

表 2-10 各组大鼠 II 型肺泡上皮细胞存活情况 ($\bar{x} \pm s$)

组 别	样 本 数	存活细胞数(A 值)	贴壁细胞数(个)	
			加 H ₂ O ₂ 前	加 H ₂ O ₂ 3h 后
NS 组	8	0.1038 \pm 0.0099	28.89 \pm 1.66	28.78 \pm 1.71
H ₂ O ₂ 组	8	0.0546 \pm 0.0004 ^a	29.26 \pm 1.84	10.20 \pm 1.30 ^b
A ₁ 组	8	0.0669 \pm 0.0039 ^{ad}	29.04 \pm 2.04	12.87 \pm 0.55 ^{bc}
A ₂ 组	8	0.0710 \pm 0.0065 ^{ade}	29.27 \pm 1.71	19.64 \pm 0.25 ^{bce}
A ₃ 组	8	0.0787 \pm 0.0092 ^{adef}	29.22 \pm 1.52	23.02 \pm 0.82 ^{abef}

表 2-11 各组大鼠 II 型肺泡上皮细胞中 HO-1 的蛋白及 mRNA 表达的比较 ($\bar{x} \pm s$)

组 别	样 本 数	HO-1 蛋白(积分 A 值)			HO-1 mRNA(Ct 值比值)		
		培养 20h	培养 40h	培养 60h	培养 20h	培养 40h	培养 60h
NS 组	8	1.25 \pm 0.13	1.24 \pm 0.10	1.25 \pm 0.11	20.92 \pm 1.25	22.33 \pm 1.67	22.99 \pm 1.95
A ₁ 组	8	1.55 \pm 0.08 ^a	1.57 \pm 0.11 ^a	1.59 \pm 0.12 ^a	22.88 \pm 1.45 ^a	23.11 \pm 2.24 ^a	24.32 \pm 1.74 ^a
A ₂ 组	8	1.57 \pm 0.12 ^a	1.56 \pm 0.10 ^a	1.60 \pm 0.09 ^a	25.65 \pm 1.59 ^{ab}	27.47 \pm 1.32 ^{ab}	30.63 \pm 2.53 ^{ab}
A ₃ 组	8	1.60 \pm 0.11 ^a	1.58 \pm 0.13 ^a	1.61 \pm 0.08 ^a	26.23 \pm 2.14 ^{abc}	29.78 \pm 2.92 ^{abc}	32.63 \pm 3.74 ^{abc}

注：NS 组：生理盐水组；H₂O₂ 组：过氧化氢损伤组；A₁ ~ A₃ 组：阿司匹林 50mg、100mg、200mg 与处理组。与 NS 组比较，^a $P < 0.05$ ；与 A₁ 组比较，^b $P < 0.05$ ；与 A₂ 组比较，^c $P < 0.05$ 。

由于本试验的试验对象为 AEC II，属纯化、原代培养的细胞，可认为细胞的生物学特性具有良好的一致性，若能严格采用随机化分组则更好。因此，从统计学的角度来看，本研究符合试验设计的基本要求。

从数据的统计学处理来看，测量指标共有 4 个，分别是存活细胞数(A 值)、贴壁细胞数(个，该指标应该是单位体积或面积的贴壁细胞计数)、HO-1 蛋白含量及 mRNA 表达。上述指标的测量结果均为定量资料，本文采用 $\bar{x} \pm s$ 来进行统计描述是否适宜，需要对原始数据进行正态性检验，如表示 mRNA 表达的指标 Ct 值的比值服从正态分布吗(一般情况下，比值很少服从正态分布)？

关于组间比较统计分析方法，本文描述为“采用 SPSS 13.0 统计软件对数据进行分析，多样本均数比较用单因素方差分析， $P < 0.05$ 为差异有统计学意义”。很显然，表 2-10 的资料中，“存活细胞数(A 值)” H₂O₂ 组的标准差为 0.0004，而 A₃ 组的标准差为 0.0092，两者相差 23 倍，可初步判断组间方差不齐，不满足方差分析的条件。指标“贴壁细胞个数”记录了加 H₂O₂ 前和加 H₂O₂ 3h 后两个时间点的数据，测量结果属于重复测量资料。但本例中的试验对象是“原代细胞”，而细胞的

生物学特性具有良好的一致性,因此可以认为加 H_2O_2 前各组之间具有较好的均衡性,此时若通过加 H_2O_2 3h 后的测量结果来反应组间处理效应的差异,当数据满足方差分析的条件(正态分布、方差齐同)时,用单因素方差分析进行处理尚可。但表 2-11 的数据是典型的重复测量数据,应按照前述的重复测量数据的统计分析方法进行处理。

【案例 2-9】 目的是探讨大鼠重度烫伤后胰岛素强化治疗拮抗心肌细胞凋亡的作用及机制。方法:将 18 只 SD 大鼠按随机数字表法分为假伤组、烫伤组和胰岛素强化治疗组(强化治疗组),每组 6 只。制作 30% 总体表面积(TBSA)Ⅲ度烫伤模型及胰岛素强化治疗模型。伤后 6 h 取大鼠左室心肌组织,采用原位末端缺刻标记法(TUNEL)检测心肌细胞凋亡,用免疫组化和蛋白质免疫印迹法(Western Blotting)分别观察 3 种凋亡相关基因天冬氨酸特异性半胱氨酸蛋白酶 3(caspase-3)、Bax 和 Bcl-2 的蛋白表达。

【案例辨析与释疑】 从上述摘要信息可知,本例属于“完全随机设计”的动物试验,试验设计的三个基本要素分别是:受试对象为“SD 大鼠”,处理因素有三个水平,分别是“假伤组、烫伤组、胰岛素强化治疗组”,试验效应为 SD 大鼠心肌组织的检测指标,分别是细胞凋亡指数(AI)、caspase-3、Bax 蛋白表达及 Bcl-2 蛋白表达;试验设计遵循“对照、重复、随机化”三项基本原则,即设立了对照组、有一定的样本含量(每组 6 只 SD 大鼠)、18 只 SD 大鼠进行了随机化分组。因此,从试验设计的角度来看,本研究基本满足试验设计的要求。

本例中,原作者将统计学处理描述为“数据以均数 \pm 标准差表示,采用 SPSS 10.0 软件进行方差分析和 t 检验, $P < 0.05$ 为差异有统计学意义”。试验结果摘录见表 2-12。

表 2-12 胰岛素强化治疗对烫伤大鼠心肌细胞 AI 及凋亡相关基因蛋白表达的影响 ($\bar{x} \pm s$)

组别	动物数	AI (%)	免疫组化法检测蛋白表达			Western Blotting 法检测蛋白表达		
			caspase-3	Bax	Bcl-2	caspase-3	Bax	Bcl-2
假伤组	6	0.6 ± 0.4	1.36 ± 0.95	2.24 ± 1.04	8.01 ± 2.56	1	1	1
烫伤组	6	13.1 ± 3.4^a	13.72 ± 4.13^a	29.64 ± 5.42^a	3.39 ± 1.52^a	5.72 ± 2.13^a	4.64 ± 1.42^a	0.69 ± 0.42^a
强化治疗组	6	6.7 ± 1.8^c	8.88 ± 3.36^b	14.43 ± 3.69^b	8.61 ± 3.72^b	2.18 ± 0.86^b	2.87 ± 1.35^c	3.57 ± 1.70^b

注:与假伤组比较, $^aP < 0.01$;与烫伤组比较, $^bP < 0.05$, $^cP < 0.01$ 。

关于统计资料的描述:从表 2-12 可以看出,四个检测指标均为定量资料,原作者对所有指标均采用 $\bar{x} \pm s$ 来描述,没有考虑资料的分布是否服从正态分布(如 AI 一般不服从正态分布),这样处理数据的方法是欠妥的。正确的做法是:对于定量资料,原作者在选择描述指标时,首先应对资料的分布类型进行检验,若资料服从正态分布,则选择 $\bar{x} \pm s$ 来描述资料的分布特征,对于不服从正态分布或分布类型未知的资料,应选择中位数和四分位数间距来进行描述。

关于假设检验方法的选择:本例属于完全随机设计三个样本定量资料的比较,因此研究者在选择假设检验方法时应该首先对资料的正态性和方差齐性进行检验,若满足正态性和方差齐性,才能用单因素方差分析进行组间差异的比较,否则应选择完全随机设计多个独立样本比较的非参检验方法(Kruskal Wallis H 检验)。本例中,原作者并未提及资料是否满足方差分析的条件(正态分布、方差齐同),对所有指标都采用方差分析进行比较是不适宜的。原作者在进行组间两两比较时采用了重复进行 t 检验,这样会增大犯 I 类错误的概率,显然是不适宜的。

2.5 本章常见的统计学问题小结

基于上述的个案分析,本节将《中国危重病急救医学》杂志中常见的统计学问题进行归纳与总结,进一步阐述应对各类常见统计学问题的方法,以便帮助读者更好地理解医学研究中的统计设计思想和统计方法的正确运用。

2.5.1 研究设计中常见的统计学问题

在临床试验设计方面,研究者首先需要分清临床试验类型是探索性临床试验,还是确证性临床试验。探索性临床试验一般样本量较小,对照设立也不严格,其试验结果仅能初步显示干预措施有一定的作用,但不能作为证实干预措施有效地证据。如果研究者需要获得干预措施有效性的可靠证据,必须通过大样本随机对照临床试验获得,也就是确证性临床试验。在《中国危重病急救医学》杂志上发表的临床试验类文章中,有些作者通常将小样本的探索性临床试验结果作为评价干预措施效果的证据,并据此推论出干预措施效果显著的结论,这显然是不合适的。如案例2-3大量脑干出血微创术的临床应用一文中,原作者仅观察了8例严重脑干出血患者应用“特殊部位脑出血微创定位器”和微创术救治后的效果,便据此得出“大量脑干出血采用微创术抽吸血肿疗效显著,不仅能显著降低病死率,而且有望明显降低致残程度”的结论,显然是不适宜的。

在临床试验设计方面,研究者不仅应该明确试验设计的三个基本要素(受试对象、处理因素、试验效应),并遵循试验设计的四个主要基本原则(对照、重复、随机、均衡),必要时,还应该考虑应用“盲法”来避免研究者和受试者的主观因素对试验结果干扰。当观察指标是一个受主观因素影响较大的变量时,建议研究者最好能够采用“盲法”来避免主观因素的影响。

观察性研究设计中存在的主要问题是样本的代表性和对照的设立。《中国危重病急救医学》杂志上发表的临床观察类文章很多,样本一般是来自一所医院特定时期的病例,当样本量较小时,样本的代表性较差。这类临床观察类文章中存在的另一个显著问题是缺乏对照。研究者往往仅对收集到的单组病例的临床资料进行汇总分析,基于此得到的结论若作为临床证据是缺乏科学性的。因此,若要得到相对科学可靠的结论,研究者需要在实践中精确地提炼科学问题,明确研究目的,通过严谨的研究设计(包括专业设计和统计设计)进行临床证据的收集,并运用适宜的统计分析方法对数据进行统计学处理,进而得到有价值的结论,为循证临床决策提供依据。

2.5.2 统计描述和统计推断中常见的统计学问题

统计描述一般包括统计指标的计算及统计结果的表达。统计指标选择应该考虑资料的数据类型和分布特征。若是定性资料,常用的相对数指标有率和构成比;若是定量资料且呈正态分布,则适宜用 $\bar{x} \pm s$ 来描述资料的分布特征,若定量资料为偏态分布或分布类型未知,则适宜用中位数描述其集中趋势,四分位数间距描述其离散趋势。《中国危重病急救医学》杂志上发表的文章在统计描述方面存在的主要问题是:定量资料通常都采用 $\bar{x} \pm s$ 来描述,而很少考虑资料是否满足正态分布要求。若统计描述指标选择不当,不仅不能很好地描述资料的分布特征,也影响后续假设检验方法的正确运用。

统计结果的表达主要是指统计图表的应用。《中国危重病急救医学》杂志上发表的文章在统计图表应用方面存在的问题不太多。统计图表的应用主要体现在图表的规范性上。通常科技期刊对论文中统计图表的应用都有明确的要求。此处不再赘述。

统计推断是指如何由样本统计量来推断总体参数,包括参数估计和假设检验。《中国危重病急救医学》杂志上发表的文章中很少有报告参数估计结果的,一般只对组间的差异进行假设检验。假设检验方法的选择非常重要。通常情况下,定性资料的组间比较多用 χ^2 检验;定量资料若满足“正态分布和方差齐同”的条件,则应选择 t 检验或方差分析进行组间比较,若不满足上述条件,则应选择非参数检验方法;等级资料的组间比较一般采用非参数检验方法。在选择假设检验方法时,除了考虑上述的一般原则外,还需要考虑研究的设计类型,尤其是一些比较复杂的研究设计类型,如重复测量设计、析因设计、正交设计等。《中国危重病急救医学》杂志上发表的文章中有许多研究属于重复测量设计,原作者在处理此类数据时很少用到重复测量资料特有的统计分析方法,包括重复测量资料的单变量方

差分析(ANOVA)或多变量方差分析(MANOVA)以及混合效应方差分析模型。多数研究者在处理此类数据通常采用方差分析或在各个时间点进行多次重复 t 检验,显然是不适宜的。

2.5.3 统计结果解释中常见的统计学问题

统计结果解释通常出现在论文的“讨论”部分,是作者针对假设检验结果,通过引用“ P 值”,对假设检验结果做出专业上的解释。在这里特别需要注意的是,“ P 值”的大小只能说明“统计学意义的显著”,不说明“实际效果的显著”。如临床试验中,某项处理在临床上有“显著”的疗效,但可能会由于临床试验的样本量过小,通过假设检验得到了较大的“ P 值”,也就是得到了统计学不显著的结论;相反,在临床上疗效“不显著”的处理,当样本量很大时,“ P 值”可能很小,从而得出“统计显著”的结论。因此,对于“ P 值”的解释一定要结合专业知识进行判断,不能将“统计显著”和“专业显著”混为一谈。

参 考 文 献

- [1] 夏锡仪, 谭玉林, 孙亚伟等. 低分子肝素预防外科术后患者的肺栓塞, 中国危重病急救医学, 2011, 23(11): 661-664.
- [2] 李游, 田伟盟, 武祺等. 中西医结合治疗 100 例老年重症肺炎的临床疗效分析. 中国危重病急救医学, 2011, 23(1): 44-45.
- [3] 耿黎明, 马飞, 许建军等. 大量脑干出血微创术的临床应用, 中国危重病急救医学, 2011, 23(10): 638-639.
- [4] 马春花. 762 例住院患者医院感染横断面调查分析. 中国危重病急救医学, 2011, 23(2): 108-109.
- [5] 张在其, 陈文标, 陈玮莹等. 广州市 97 823 例院前急救患者流行病学分析, 中国危重病急救医学, 2011, 23(2): 99-103.
- [6] 刘晖, 马清艳, 孙峰等. 肺血栓栓塞症 252 例临床诊治体会, 中国危重病急救医学, 2012, 24(7): 398-401.
- [7] 史东升, 周静敏, 马淑萍, 12 例心力衰竭致大量胸腔积液患者临床分析, 中国危重病急救医学, 2011, 23(6): 332.
- [8] 王琼康, 黄晓莲, 张云强等. 50 例急性肺栓塞临床诊治分析, 中国危重病急救医学, 2011, 23(1): 46-47.
- [9] 李嘉嘉, 尚云波, 熊俊光等. 机械通气下内镜治疗食管胃底静脉曲张破裂出血的救治体会, 中国危重病急救医学, 2012, 24(2): 95.
- [10] 王传海, 童朝晖, 詹庆体等. 外膜肺氧合治疗急性呼吸窘迫综合征的临床疗效观察, 中国危重病急救医学, 2012, 24(2): 83-84.
- [11] 巩传勇, 傅强, 杜超. 性激素对腹部外科疾病伴全身炎症反应综合征患者预后的影响, 中国危重病急救医学, 2011, 23(11): 673-676.
- [12] 张铮, 沈华秦, 海东等. N-乙酰半胱氨酸对大鼠肺缺血/再灌注损伤诱导细胞凋亡的保护机制, 中国危重病急救医学, 2012, 24(2): 111-115.
- [13] 钱明江, 陈森, 王洪敏等. 阿司匹林预处理对原代培养大鼠 II 型肺泡上皮细胞抗氧化损伤的影响研究, 中国危重病急救医学, 2011, 23(4): 240-250.
- [14] 吕根法, 石宏伟, 樊磊等. 胰岛素强化治疗拮抗烫伤大鼠心肌细胞凋亡的机制研究. 中国危重病急救医学, 2011; 23(12): 714-717.
- [15] 孙振球. 医学统计学(第2版). 北京: 人民卫生出版社, 2006.

第3章 医学综合

《中华医学杂志》为中华医学会会刊，是综合性医学学术期刊，国内外公开发行。该刊主要全面反映我国医学科研成果、快速传递世界前沿信息、推广现代先进技术、及时交流防病治病经验、大力普及医学科技新知。《中华医学杂志》设有述评、医学动态、医药卫生策略探讨、政策法规透视、论著、临床病理讨论、疑难病例析评、学术争鸣、循证医学、病例报告、技术交流、临床医学影像、药物与临床、标准与规范、会议纪要、专题笔谈、综述、讲座、继续教育园地、国内外学术动态、读者来信、书评(或书讯)、人才交流与招聘及学术活动预告等栏目。《中华医学杂志》等中华系列学术刊物对中国医学研究的发展起到了推动作用。

本章结合《中华医学杂志》的文献中的一些具体实例，从统计学角度，为读者介绍如何在医学研究中正确运用统计思想、理论与方法，提高识别“科研设计、统计表达与描述、统计分析”三方面出现错误的能力及其正确实施的理论指导和技术支持。

3.1 《中华医学杂志》论文分析

《中华医学杂志》发表论文的涵盖范围广泛，围绕中国医学科研攻关课题和严重危害广大人民健康的常见病、多发病的防治，反映中国医学最新的科研成果，理论与实践结合，积极推广医药卫生领域的新技术、新成果，及时交流防病治病的新经验，普及医学科技新知识，为提高中国广大医药卫生人员的思想和职业道德修养、医学科学理论和业务水平，推动中国医学科技进步和知识创新，以及繁荣中国医学科技出版事业，发挥了重要作用。

中华医学杂志为月刊时，每年收稿 2000 篇左右，刊用率约为 15% 左右。2001 年改为半月刊以后，每年收稿约 4000 篇左右，刊用率约为 20% 左右。2005 年改为周刊后，在 2005 年全年收稿 4280 篇，刊用率为 28%。中华医学杂志实行审稿制。中华医学杂志发表的论著类文章中，90% 左右是医学基础研究和临床医学各专业具有国内领先水平的科研成果，部分达到国际先进水平。论著类文章 80% 以上都受国家和省部级基金资助，其中 70% 以上都受国家级基金资助。

《中华医学杂志》的撰稿要求中，涉及研究的统计设计、统计描述与统计分析等的具体要求。其中，针对研究设计的要求有：当研究对象为人时，作者应该说明是否符合人体试验伦理学标准，并得到伦理委员会的批准是否有患者知情同意。调查设计应交代是前瞻性、回顾性还是横断面调查研究；试验设计应交代具体的设计类型，如属于自身配对设计、成组设计、交叉设计、析因设计或正交设计等；临床试验设计应交代属于第几期临床试验、采用了何种盲法措施、受试对象的纳入和剔除标准等。应交代如何控制重要的非试验因素的干扰和影响。

针对统计学的要求有：应写明所用统计分析方法的具体名称(如成组设计资料的 t 检验、两因素析因设计资料的方差分析等)和统计量的具体值(如 $t=3.45$)，并尽可能给出具体的 P 值(如 $P=0.023$)；当涉及总体参数时，在给出显著性检验结果的同时，再给出 95% 可信区间。对于服从偏态分布的定量资料，应采用 $M(Q_1 \sim Q_3)$ 方式表达，不宜采用 $\bar{x} \pm s$ 方式表达。对于定量资料和定性资料，应根据所采用的设计类型、资料所具备的条件和分析目的，选用合适的统计分析方法，前者不应盲目套用 t 检验和单因素方差分析，后者不应盲目套用卡方检验。要避免用直线回归方程描述有

明显曲线变化趋势的资料。不宜用相关分析说明两种检测方法之间吻合程度的高低。对于多因素、多指标资料,要在一元分析的基础上,尽可能运用多元统计分析方法,以便对因素之间的交互作用和多指标之间的内在联系做出全面、合理的解释。使用相对数时,分母不宜小于20;要注意区分百分率与百分比。统计学符号按《统计学名词及符号》(GB 3358—1982)的有关规定书写,一律用斜体。

针对图表表达的要求有:每幅图、表应有简明的题目;要合理安排表的纵、横标目,并将数据的含义表达清楚;表内数据要求同一指标保留的小数位数相同,一般比可准确测量的精度多一位。图不宜过大,高与宽的比例应掌握在5:7左右。图的类型应与资料性质匹配,并使数轴上刻度值的标法符合数学原则。

3.2 统计设计中的常见问题及案例辨析

3.2.1 研究设计的类型、要素与原则

【案例3-1】 比较快速人工心脏起搏与硝普钠在胸主动脉腔内修复治疗中的安全性、有效性及对移植物流定位准确性的影响。研究入选2007年9月至2009年2月在广东省人民医院心内科因胸主动脉疾病择期行胸主动脉腔内修复治疗(TEVAR)的197例患者(夹层175例,动脉瘤22例)。随机分为硝普钠组($n=98$)和快速人工心脏起搏组($n=99$),在腔内修复治疗术中分别应用上述两种方法以助移植物的准确定位与释放。比较两组的血流动力学指标、移植物流定位及释放准确性、手术时间、手术前后肾功能和神经认识功能变化以及内漏和截/偏瘫发生率。

请读者结合试验设计相关知识分析该论文属于什么研究类型?试验设计的三要素是什么?该试验研究采用了何种对照?

【案例辨析】 这是一个临床随机对照试验,试验因素是胸主动脉腔内修复治疗方法(它的两个水平分别为快速人工心脏起搏器与硝普钠药物),重要的非试验因素可能有:患病的严重程度、年龄、身体素质等;受试对象为因胸主动脉疾病择期行胸主动脉腔内修复治疗的患者;试验效应是血流动力学指标、移植物流定位及释放准确性、手术时间、手术前后肾功能和神经认识功能变化,以及内漏和截、偏瘫发生率等。

该试验采用现有的常规方法作为对照。这种采用现有的标准方法、常规方法、标准值或参考值作为对照的方法,称为标准对照,是临床试验中常用的对照设置方法之一。未交代清楚采用了何种随机分组的方法,若采用完全随机分组,“夹层”与“动脉瘤”两种状态的患者很难在两种治疗方法组中达到十分理想的均衡,应按这两种状态进行分层随机化。

由于研究局限于一所医院的单中心,因此结果的外延有限。

【案例释疑】 医学研究的类型很多,不同的分类标准有不同的研究类型。医学研究设计从研究目的可以分为探索性研究与验证性研究;从考察因素多少可以分为单因素研究与多因素研究;从研究的对象可分为以正常人群为基础的调查研究或社区干预试验、以患者为基础的临床试验和以动物或其他试验材料为基础的动物试验;从研究的时限来看,可以分为前瞻性研究、回顾性研究和横断面研究;根据医学研究过程的不同发展阶段,联合国教科文组织又把医学科研究分为三大类,即基础研究、应用研究、发展研究;从研究是否对研究对象施加干预,可以分为试验性研究与观察性研究。

试验性研究是对研究对象人为施加干预的研究。研究者通过随机分组,对不同处理组的研究对象施加不同的干预,从而比较不同干预措施间的效果。

观察性研究是对研究对象不加任何干预措施,在完全“自然状态”下观察、记录各研究对象的特征,描述现象的本质,比较不同环境条件下疾病的发生、发展。

观察性研究与试验性研究的最大区别是：不能对研究对象进行完全随机分组，不能按照研究者的主观意愿对研究对象进行干预，只能被动观察；观察性研究中的随机化多指随机抽样，试验性研究中的随机化多指随机分组。

1. 试验设计概述

研究者在确定了研究内容后，如何科学合理地进行科学研究，得到可信的结论？在科研过程中应注意哪些环节呢？一个好的科学研究，首先应具有一个科学完善的科研设计方案，包括专业设计和统计研究设计。统计研究设计根据研究类型分为试验设计、临床试验设计和调查设计。本例属于临床试验研究，应首先进行临床试验设计。

试验设计的核心内容是试验设计的“三要素”、“四原则”和“设计类型”；而临床试验设计应基于试验设计，还应考虑伦理道德、采取盲法等技术消除来自受试者和研究者心理和人为因素对试验结果的干扰。三要素分别是受试对象、试验因素和试验效应。四原则是指随机、对照、重复与均衡。设计类型是由试验因素及其水平按某种规则组合起来的一种架构，需根据试验因素个数、水平数、试验因素之间的关系和有关专业知识等综合起来方可确定。

2. 试验设计的三要素

试验研究由三个基本要素组成，即影响因素（包括试验因素和重要非试验因素）、受试对象和试验效应。如用某种传统西药或中成药治疗缺铁性贫血病人，观察比较两组病人血红蛋白的上升趋势，该研究中所用的两种药物称为处理因素，缺铁性贫血病人称为受试对象，血红蛋白称为观测指标，其取值大小就是试验效应。在医学研究中，如何正确处置三要素是科研中专业设计的关键问题。

(1) 受试对象。受试对象是试验因素的承受者。受试对象的选择应以试验研究的目的为依据，并结合基本常识和专业知识的确定。在临床试验中受试对象通常为患者，应注意其性别、年龄、病情和病程，民族、职业、文化程度和经济状况，这些往往都是对观测结果可能造成不可忽视影响的重要非试验因素，应设法确保它们在不同处理组之间的平衡性。需遵循以下原则：①每项临床试验研究必须通过有资质伦理委员会的审批，受试者应自愿同意参加试验，并签署知情同意书，以使此项临床试验研究符合伦理道德。②诊断明确。对于特定的疾病，研究者应该有统一的诊断标准，最好是国际公认的标准，或被国内同行一致认可的标准；若没有统一的标准，可以自己设定且得到同行专家的一致认可，必须尽可能地采用客观的标准，在操作时能够明确诊断。③依从性好。患者由于心理、社会、经济等多方面原因而可能出现忘记服药、中途退出试验或出现换组。其次，由于病情急剧恶化或存在难以忍受的副作用，必须中途退出试验，这些不依从性表现必然干扰试验计划的实施。因此，必须充分关心体贴病人，做好思想工作，使病人建立充分信任与依赖医务人员的心理状态，从而提高依从性。并且应当控制试验时间，试验时间过长往往依从性降低。④被选择的对象应该从试验研究中可能受益（通常应利大于弊）。⑤已知试验对其有害的人群，不应作为试验对象；例如，有消化道出血史者不应作为抗炎药物的试验对象。⑥对于一些研究对象患有可能会影响试验结果的疾病，或这些疾病本身并不影响试验结果，但治疗这些疾病所用的药物或措施可能影响试验结果，对于这些病例必须被排除。此外，研究者还要结合实际情况制定更详细的纳入、排除和剔除标准。

(2) 影响因素。所谓影响因素，就是可以明显改变结果变量取值的性质相同的若干个试验条件的总称。影响因素包括试验因素和非试验因素。所谓试验因素，就是在试验中，研究者希望着重考察的某些性质相同的试验条件（特殊情况下是事物的两个对立方面）。为了尽可能减少各组受试对象的自身条件或外界环境或研究者主观意识等（统称为非试验因素）对试验结果的影响，在试验设计时，研究者应根据专业知识确定几个重要的非试验因素，对重要的非试验因素作有计划的安排，

使它们对试验因素不同水平组的影响达到均衡一致,同时,也便于在统计分析时,将其作用排除,从而更准确地评价试验因素对试验结果的影响。需要注意的是,任何试验效应都是多种因素综合作用的结果,我们不可能也没有必要在一次或几次研究中,研究所想到的一切有关因素,只能抓住对试验结果影响较大的主要因素加以研究。

(3) 试验效应。试验效应是试验因素作用于受试对象的反应,是研究结果的最终体现,也是试验研究的核心内容。它是通过具体的观测指标来实现的,即观测指标是用来反映试验因素作用强弱的重要“尺子”,必须结合试验因素的性质和特点、仪器、试剂和技术水平等多方面综合考虑。这些指标包括定性指标和定量指标,也包括主观指标和客观指标。选择指标的基本要求是:指标应具有客观性、特异性、灵敏性和精确性。如果指标选择不当,未能准确反应试验因素的作用,获得的研究结果就缺乏科学性,因此,选择好的指标是关系研究成败的重要环节。选好试验效应的指标以后,还要规定指标观察的常规方法,如观察方法、标准、时间、记录方法及记录格式等。指标的观察或测量应避免偏性,否则会影响结果的比较和分析。研究者可以通过培训来掌握统一评价标准,或是在试验设计时常采用盲法,都可以消除这种偏性。

试验设计是以 R. A. Fisher 为主在 20 世纪 30 年代发展起来的。所谓试验设计,就是将处理分配给试验单位的方法。好的试验设计应能较好地控制随机误差,避免或减少非随机误差,以较少的试验对象取得较多而可靠的信息,达到经济、高效之目的。为此,试验设计必须遵循如下四个基本的统计学原则。

1) 对照原则

有比较才能鉴别。设立对照的目的是为了体现某个或某些特定试验条件的效应。所设立的对照组必须与试验组之间在一切重要非试验因素方面达到均衡一致。对照组设立后,应对各对比组的基线情况进行比较,以检验其均衡性。

对照原则的作用是提高试验的鉴别能力,将处理因素给受试者带来的效应(症状、体征或指标的改变)与其他因素(如疾病的自然进展、观察者或患者的期望、其他治疗措施等非处理因素)造成的效应区分开来。对照有下列几种形式。

(1) 安慰剂对照。在临床研究中,设置安慰剂对照可用于消除受试者心理因素导致的偏倚,还可以消除疾病自然进程的影响。值得注意的是,安慰剂的使用需谨慎,不能损害受试者的健康,多用于尚无有效药物治疗的疾病。有时从医学伦理学的角度考虑,可以在每个受试者都给予一种标准基础治疗的基础上,试验组再给予试验药物,对照组再给予安慰剂。

(2) 空白对照。对照组不接受任何具有实质性的处理。在使用空白对照时,应注意不要违背伦理道德。

(3) 标准对照。用现有标准方法、常规方法、标准值或参考值作为对照。试验研究中常用于某种新方法是否能代替传统方法。

(4) 自身对照。自身对照是指对照与试验在同一受试对象身上进行。例如,为测量某电子血压计的准确性,以若干个人为研究对象,每人先后用精确的血压计和电子血压计测量血压,计算测量的差值,进行统计学分析。

(5) 试验对照。对照组不施加处理因素的真实水平。如研究饲料中某种营养成分的价值,在这个试验中,在基础饲料基础上加上某种营养成分,则该营养成分是处理因素(是否含有此营养成分)的真实水平,而基础饲料则是试验组与对照组的“本底”或称“载体”,两组除是否含有这种营养成分外,其余条件都一致,这样才能分析这种营养成分存在与不存在之间的差异。因此,当处理因素的施加需伴随其他“载体”(本例为“饲料”),而这种“载体”可能影响试验结果时,应设立试验对照(即仅有“载体”),以保证组间的均衡性。

(6) 历史对照。将研究者以往的研究结果或文献上他人的研究结果与本次研究结果作对照。该对照一般用于考核时间因素带来的变化,或用于公认的难治性疾病,除恶性肿瘤等以外,一般不宜使用。因为随着时间的推移,医学的进步,历史资料往往不具有可比性。

2) 随机原则

随机是为了保证各处理组间在大量不可控制的非处理因素的分布方面保持均衡可比而采取的一种统计学措施。因此,在试验对象的抽样、分组以及试验实施过程中均应遵循随机化原则。随机应体现在如下几个方面。

(1) 抽样随机:每一个符合条件的研究对象参加研究的机会相同,即总体中每个个体有相同的会被抽到样本中来,从而保证所得到的样本具有代表性,以使研究结论具有普遍意义。

(2) 分组随机:每个试验对象分配到不同处理组的机会相同,从而保证各处理组间试验对象尽可能均衡一致,以提高组间的可比性。

(3) 试验顺序的随机:每个研究对象先后接受处理的机会相同。是为了平衡试验顺序对结果的可能影响。

分组随机化是指在试验分组时,每个观察对象具有相同的机会(概率)被分配到试验组或对照组中去。分组随机化是试验设计中多个步骤中的一个步骤,是给受试对象分组的一种方法,关键在于消除人为因素对分组造成的影响,不能理解为随心所欲的“随便”、“随意”和“非选择性”。其主要作用如下。

(1) 提高试验组和对照组在各种已知或未知特征(如患者的年龄、性别、病情、用药史等)等方面的均衡性,或者说,在随机分组后、施加干预以前,各组观察对象除处理因素取不同水平外,受其他因素的影响是几乎相同的。施加干预以后如果发现试验组和对照组之间有差异,则认为干预有效。

(2) 避免因研究者主观愿望破坏试验组和对照组的均衡性,如有意(或无意)将轻病人分配到试验组,将重病人分配到对照组。因此,在临床新药试验中,随机化分组过程和分组结果必须作为档案保存,以便药物监督部门核查。

(3) 随机化分组是所有统计假设检验(如 t 检验、 F 检验)推论因果关系的基础。

3) 重复原则

重复是指在相同试验条件下进行多次研究或多次观察,以提高试验的可靠性与科学性。每个特定试验条件下重复多少次试验合适?属于样本含量估计问题,应根据研究目的、研究和设计类型、现有的人力物力和财力、某些已知的前提条件和对结果精确度的要求等,找到相应的估计样本含量的公式计算出来为宜。

4) 均衡原则

均衡是指各对比组之间除处理因素取不同水平外,其他重要的、可控制和不可控制的非处理因素的分布尽量保持一致。例如,不同对比组试验对象在性别构成、平均年龄及平均健康状况上应保持一致。在临床试验中,还应考虑到病情、病程以及过去接受治疗的情况等在不同组间的均衡性。

3.2.2 研究对象的选择与代表性

【案例 3-2】 分析中国年轻乳腺癌的临床病理特征,并探讨年轻乳腺癌患者的预后。回顾分析北京肿瘤医院乳腺中心 1994 年 12 月至 2003 年 12 月收治的 1538 例 I ~ III 期可手术原发性乳腺癌患者的临床资料,其中年龄 ≤ 60 岁且有完整随访资料者 1075 例。按年龄将 1075 例患者分为年轻组(≤ 40 岁, 208 例)和对照组(41 ~ 60 岁, 867 例),分析两组患者的预后及临床病理特征之间的差异。

请读者结合试验设计相关知识分析该论文属于什么研究类型? 如何选取研究对象合适?

【案例辨析】 本研究采用的是回顾性的观察性研究。通过与中老年乳腺癌患者的比较来探讨年轻乳腺癌患者的特征。研究对象来源于一所医院一定时间段内的所有可手术原发性乳腺癌患者。由于研究对象选取的局限性, 结果只代表该地区或该医院范围, 而不应该拓展到全中国范围。如果要分析中国年轻乳腺癌的临床病理特征, 那么在选取研究对象时应该在全国范围内随机抽样, 这样才具有代表性。

【案例释疑】 医学实践与研究中, 要直接研究无限总体通常是不可能的, 即使是有限总体, 由于人力、物力、时间、条件等限制, 要对其中每个观察单位进行研究观察, 有时也是不可能的, 也不必要。而只是从总体中随机抽取部分观察单位, 其变量实测值构成样本, 目的是用样本指标推断总体特征。这种推断需要经过严谨的试验设计, 以样本的可靠性和代表性为基础。样本的可靠性: 主要是使样本中每一观察单位确属同质总体, 如临床试验研究中, 需要有正确的诊断, 对病种、病程、病期、病型等要严格划分; 动物试验中, 应规定动物种属、品系、窝别、性别、体重及生理状态力求一致, 以排除和控制非试验因素的干扰。

样本的代表性由随机抽样保证, 随机抽样指总体中的每个观察单位都有同样的机会被选入样本, 这样获得的样本称为随机样本。常用的随机抽样方法有以下4种。

(1) 简单随机抽样。简单随机抽样是在总体中以完全随机的方法抽取一部分个体组成样本的抽样方法, 即总体中每一个观察单位都有同等机会被抽取进入样本。其优点是均数(或率)及标准误差的计算较简便, 但对有限总体和无限总体标准误差的计算公式有所不同。其缺点是当总体例数较多时, 要对观察单位逐一编号较麻烦。

(2) 系统抽样。系统抽样又称为等距抽样或机械抽样, 是指随机地在所要抽样的有序名单中每隔若干个个体抽取一个个体的抽样方法。这种方法的优点是简便易行, 容易得到一个按比例分配的样本, 但并没有专门的计算系统抽样的抽样误差公式, 系统抽样的抽样误差与简单随机抽样相仿, 一般都借用简单随机抽样时抽样误差的计算公式。系统抽样的缺点是: 当总体的已有的顺序号有周期趋势或规律性时, 系统抽样将产生明显的偏性。

(3) 分层抽样。按照与研究目的明显有关的因素(或某种特征), 将总体分为若干类型或区域, 统计上叫“层”, 然后从每一层内按比例抽取一定数量的观察单位, 将各层的观察单位合起来组成样本, 这种抽样方法称为分层抽样。

(4) 整群抽样。首先将总体按照某种与研究目的无关的分布特征(如地区范围; 不同的团体; 病历、格子等)划分为若干个“群”组, 每个群包括若干观察单位; 然后根据需要随机抽取其中部分“群”, 并调查被抽各“群”中的全部观察单位。

整群抽样的优点是便于操作, 容易控制调查质量。简单随机抽样需要对调查总体的全部观察单位进行统一编号。在实际工作中, 尤其当总体范围较大时, 做到这一点往往比较困难。而总体中的地区、集团等分布特征, 则是范围清楚的、可利用的调查“群”, 故整群抽样颇为常用。其缺点是由于所得样本在总体中的分布并不均匀, 所以, 当样本例数一定时, 其抽样误差一般大于简单随机抽样的误差。抽取的“群”数越少, “群”间差异越大, 抽样误差也越大。

3.2.3 研究设计类型应用

【案例3-3】 对64层螺旋CT血管造影(CTA)诊断椎动脉狭窄的文献进行荟萃分析, 评价64层螺旋CTA对椎动脉狭窄的诊断价值。以64层螺旋CT、血管造影、椎动脉狭窄等为检索词, 检索Cochrane图书馆、PubMed、EBSCO、中国生物医学文献数据库、清华同方全文数据库等的中文和英文文献。对符合纳入标准的文献进行质量评估、异质性检验。汇总敏感性和特异性, 通过汇总受试

者工作特征曲线(SROC)和曲线下面积(AUC)以及诊断优势比(DROC)评价64层螺旋CT对椎动脉狭窄的诊断价值。

检索及文献评价:由两名评价员独立检索和评价,遇到不一致情况时经讨论确定或经第三名医师评价。文献等级评价参照循证医学指南中的分级标准进行:①独立;②研究不具有相同的征象(如性别、年龄、基础疾病及疾病的严重程度);③盲法评价结果;④检查及图像处理质量控制合理。均符合以上标准的评为A级,若其中一项标准不符合或不清楚则评为B级。

数据分析:采用Meta-disc version 1.4软件进行统计学处理。对资料进行异质性检验;汇总敏感性、特异性;计算诊断优势比(diagnosis OR, DOR)和汇总受试者工作特征曲线(SROC)面积(AUC)。

请读者结合相关知识分析该论文属于什么研究类型?研究有何不妥之处?

【案例辨析】该论文是对相关研究的二次研究,即系统评价与Meta分析。

本论文在检索及文献评价中,“文献等级评价参照循证医学指南中的分级标准”列出四项标准,但是第四项显然是作者自己加上去的,建议放在研究纳入标准中。另外,作者参考的是2000年的书籍;而实际上文献评价标准已经更改,且不建议采用量化等级来标注文献评价结果,目前Cochrane推荐用于证据质量评价及推荐强度评级的GRADE(Grading of Recommendations Assessment、Development and Evaluation)方法来评价研究质量。

数据分析中,原作者并没有分析研究偏倚。系统评价与Meta分析需要详细评估可能影响数据综合结果的可能存在的偏倚,如发表偏倚和研究中的选择性报告偏倚,并应该在讨论中分析研究、结局和整个系统评价的局限性。原作者在结果中也没有阐述文献纳入的具体过程,建议增加各阶段信息收集的流程图来直观表述。

【案例释疑】研究设计的恰当性是评估证据质量的第一步。根据研究资料的来源,研究设计可以分为原始研究和二次研究。原始研究的结果直接来源于病人或人群,如随机对照试验和队列研究;二次研究是基于原始研究的研究,主要指各种原始研究的综述性文章,如随机对照试验的系统综述(系统评价)。

Meta分析是用统计学方法对收集的多个研究资料进行分析和概括,以提供量化的平均效果来回答研究的问题。其优点是通过增大样本含量来增加结论的可信度,解决研究结果的不一致性。Meta分析是对同一课题的多项独立研究的结果进行系统的、定量的综合性分析。它是文献的量化综述,是以同一课题的多项独立研究的结果为研究对象,在严格设计的基础上,运用适当的统计学方法对多个研究结果进行系统、客观、定量的综合分析。

系统评价与Meta分析的步骤:提出问题,查找和选择研究,研究的质量评价,收集数据,分析和表达结果,解释结果,完善和更新系统评价。

随着系统评价越来越多地用于医疗卫生决策,评估它们的方法学质量也变得越来越重要。尽管系统评价比任何单一研究都可靠,但是,系统评价的质量有高有低,决策时必须对其质量进行严格评估。评估系统评价质量时,可以参考系统评价和Meta分析优先报告的条目:PRISMA(Preferred Reporting Items for Systematic Reviews and Meta-Analysis)声明,它包含27项条目的清单和一个分四阶段的流程图。具体内容参见PRISMA声明网站(<http://www.prisma-statement.org>)。

3.3 统计分析中的常见问题及案例辨析

3.3.1 资料统计分析方法的合理选用

【案例3-4】观察不同营养支持模式对于重症急性胰腺炎的血液流变学、炎症因子水平的影响,探索重症急性胰腺炎(SAP)营养支持最佳模式。将试验动物按数字随机法分为3组:早期肠内

营养组(EEN), SAP 后 1d 开始肠内营养(EN); 过渡肠内营养组(CEN), SAP 制模后 1、2d 静脉营养(PN), 3d 半量 PN、半量 EN, 4d 后全量 EN; 肠外营养组, 术后 1d 开始全程 PN; 每组 20 只。在制模后 1、3、7d 分别对试验动物进行血液流变学检查, 制模后 1、7d 进行炎症因子[白细胞介素(IL)-6、IL-8、肿瘤坏死因子(TNF)]检测。试验结果以 $\bar{x} \pm s$ 表示。组间比较用方差分析; 同组间比较用配伍方差分析, 两两计数资料对比运用 t 检验。应用统计学软件 SPSS 16.0 进行统计, $P < 0.05$ 为差异有统计学意义。

请读者结合统计分析方法相关知识分析该论文在统计分析方面存在哪些问题? 应如何改进?

【案例辨析】 原作者采用“方差分析”、“配伍方差分析”方法分析数据。由于本研究分析的试验因素有两个: 三种不同营养支持模式和时间因素。因此, 统计分析不能单纯考虑三种不同营养支持模式, 而忽略时间因素的影响。文中对结果指标的观测是对同一批试验动物不同时间点的重复测量, 因此, 最好采用重复测量设计定量资料的方差分析方法来分析数据。

【案例释疑】

1) 资料类型

统计分析的对象是统计资料, 因此, 准确识别统计资料类型是合理选用统计分析方法的首要前提。资料类型可以分为定量资料和定性资料(名义资料和等级资料)两类。测定每个观察单位某项指标量的大小, 所得的资料称为定量资料。例如, 正常成年男子身高(cm)、体重(kg)、血红蛋白(g/L)、每分钟脉搏次数等资料。每个个体的测定结果仅仅是一个符号或状态, 所有符号或状态之间没有数量大小之分或真正意义上的先后顺序之分, 就称为定性资料或名义资料, 如某班级学生按 ABO 血型系统划分, 所得结果为 A 型、B 型、AB 型、O 型血。在类似名义资料的定义中, 若所有符号或状态之间有数量大小之分或严格的顺序之分, 就称为等级或有序资料。例如, 某种治疗方法医治 100 名某病患者, 每位患者的结果将是治愈、显效、好转、无效、死亡之一。

2) 资料统计分析方法的合理选用策略

常用统计分析方法有定量资料的统计分析方法(如单组、配对、成组设计定量资料的 t 检验、符号秩检验及秩和检验; 单因素多水平设计定量资料的方差分析及秩和检验; 各种多因素设计定量资料的方差分析; 随机区组设计定量资料的秩和检验)、定性资料的统计分析方法(各种 2×2 列联表资料的统计分析、各种 $R \times C$ 列联表资料的统计分析和各种高维列联表资料的统计分析)和简单线性相关与回归分析方法。

如本例所述, 资料满足独立性、正态性和方差齐性时, 应选用参数检验法; 不满足时, 可寻找合适的变量变换方法, 但对变换后的数据仍应检查其前提条件, 满足时, 可对变换后的数据采用参数法; 反之, 需要用相应的“非参数”法, 通常采用相应设计定量资料秩和检验。

3.3.2 误用 t 检验分析多因素设计资料

【案例 3-5】 探讨香烟烟雾暴露肺气肿小鼠肺实质中 Th17 细胞的变化及其可能机制。将 40 只雄性 Balb/c 小鼠按随机数字表法分为 4 组: 对照 12 周组、对照 24 周组、烟雾暴露 12 周组、烟雾暴露 24 周组, 每组 10 只。香烟烟雾暴露法建立小鼠肺气肿模型。HE 染色观察小鼠肺组织的改变, 计算平均内衬间隔和肺泡破坏指数; 流式细胞术检测小鼠肺实质中 Th17 细胞、Th1 细胞、Th17/Th1 细胞、CIM IL-17 趋化因子受体(CCR)6 T 细胞占 CD 4 细胞比例及 CCR6 Th17 细胞占 Th17 细胞比例; 酶联免疫吸附法(ELISA)检测小鼠肺实质中的白细胞介素(IL)-1B、IL-6、IL-23、转化生长因子(TGF)-B、 γ 干扰素(IFN)及趋化因子(CCL)20 水平, 并分析其相互关系。统计学方法: 采用 SPSS 16.0 软件进行统计学分析。正态分布的计量资料以 $\bar{x} \pm s$ 表示; 组间比较采用 t 检

验，相关性检验采用 Pearson 相关分析；以 $P < 0.05$ 为差异有统计学意义。请辨析原作者分析此资料时有无不妥之处。

表 3-1 各组小鼠肺实质流式细胞术检测结果(%, $\bar{x} \pm s$, $n=10$)

组 别	Th17 细胞	Th1 细胞	Th17/ Th1 细胞
对照 12 周组	1.80 ± 0.75	3.75 ± 1.72	0.27 ± 0.12
对照 24 周组	1.99 ± 0.59	4.16 ± 1.32	0.28 ± 0.11
烟雾暴露 12 周组	3.27 ± 1.12 ^a	10.02 ± 3.68 ^a	0.61 ± 0.30 ^a
烟雾暴露 24 周组	7.19 ± 2.24 ^{bc}	26.21 ± 6.04 ^{bc}	1.82 ± 0.52 ^{bc}

注：CCR 为趋化因子受体；^a与对照 12 周组比较， $P < 0.05$ ；^b与对照 24 周组比较， $P < 0.05$ ；^c与烟雾暴露 12 周组比较， $P < 0.05$ 。

【案例辨析】 原作者采用“ t 检验”方法比较组间差别不妥。从列表可见，本研究涉及四个组。多组均数间的比较，由于多次运用 t 检验会增大犯假阳性错误的概率，从而导致分析结果错误，因此不宜采用 t 检验方法进行比较。正确的统计分析方法应该是在满足独立性、正态性和方差齐性的前提下，可以采用相应设计定量资料的方差分析方法。那么，本例的设计类型是什么？因试验因素有两个，即是否暴露烟雾（对照与烟雾暴露）和暴露时间（12 周与 24 周），显然这两个试验因素在试验中同时施加于受试对象，假定它们在专业上地位平等，则此试验的设计类型可称为两因素析因设计。因此，可以采用两因素析因设计定量资料的方差分析方法来分析。

【案例释疑】 方差分析能够推断多个样本所来自的总体及其总体均数是否相等，但不能推断哪些总体均数之间有差别，哪些总体均数之间无差别，若用两总体均数比较的 t 检验（或 u 检验）对多个总体均数两两进行检验，则会增大犯第一类错误的概率，特别是两两比较的次数较多时。例如，六个总体均数做两两比较时，若用 t 检验两两比较，按排列组合原理则需比较 15 次，若检验水准 α 每次均取 0.05，则每次比较不犯第一类错误的概率为 $(1 - 0.05)$ ，15 次比较都不犯第一类错误的概率为 $(1 - 0.05)^{15} = 0.4633$ ，此时犯第一类错误的概率不再是 0.05，而是 $1 - (1 - 0.05)^{15} = 0.5367$ 了，即检验水准 α 变为 0.5367，远大于 0.05。因此多个总体均数间的两两比较，不宜用 t 检验（或 u 检验）。多个总体均数两两之间的正规比较方法很多，常用的有 SNK 检验（即 q 检验）和 LSD 检验等。

研究设计按试验因素（干预因素）多少可分为单试验因素设计和多试验因素设计。单试验因素设计只有一个试验因素，通常采用完全随机设计、随机区组设计和拉丁方设计等。多试验因素设计一般安排两个或两个以上试验因素，如析因设计、交叉设计等，此类设计试验因素间的交互作用往往是分析的重要内容。析因试验设计又称为完全交叉分组试验设计，它不仅可检验每一因素各水平之间的效应差异，而且可检验各因素之间的交互作用效应大小。

3.3.3 误用单因素方差分析处理两因素资料

【案例 3-6】 本文研究不同温度与持续时间下射频热凝对兔半月神经节的组织学影响。利用 R-2000B 型温控热凝射频仪和外裸长 50mm、直径 0.4mm 热敏电极，对 80 只兔半月神经节分别进行 60℃、70℃ 和 80℃ 温控热凝毁损试验，于术后第 1、7、14 和 28 天取出受损的半月神经节，HE 染色阳性者，按石蜡切片热修复抗原试验程序行免疫组化检测，观察毁损后神经丝蛋白（NF200）、乙酰胆碱酯酶（ACHE）及毒蕈碱乙酰胆碱酯酶受体（M 受体）表达情况。统计学方法：应用 SPSS 13.0 统计软件进行统计学分析，试验数据以 $\bar{x} \pm s$ 表示，组内比较采用重复测量数据方差分析，组间比较采用单因素方差分析，以 $P < 0.05$ 为差异有统计学意义。结果：同一温度下，射频持续时间相同时，随术后时间的增加，NF200、ACHE、M 受体三项指标的阳性表达率逐渐增高；而射频持续 1min 与 3min 比较，三项指标阳性表达率差异均无统计学意义（均 $P > 0.05$ ）。射频持续时间相同时，

70℃与60℃比较,三项指标阳性表达率均明显降低,差异均有统计学意义(均 $P < 0.05$);而70℃与80℃比较,三项指标阳性表达率差异均无统计学意义(均 $P > 0.05$)。结论:当射频毁损温度为70℃左右时,对半月神经节杀伤力最大,神经活性明显降低;这可能是半月神经节毁损的较适宜的温度。

【案例辨析】 本研究为动物试验,目的是研究不同温度与持续时间下射频热凝对兔半月神经节的组织学影响。温度与持续时间为两个试验因素;其中,温度因素有三个水平:60、70和80℃;持续时间因素有四个水平:术后第1、7、14和28天。由于在不同时间点观测不同个体的受损的半月神经节,各时间点不是对应同一个个体,因此,是一个非重复测量数据的某种两因素设计资料。

原文结论为“当射频毁损温度为70℃左右时,对半月神经节杀伤力最大,神经活性明显降低;这可能是半月神经节毁损的较适宜的温度”;然而,结果显示“70℃与80℃比较,三项指标阳性表达率差异均无统计学意义(均 $P > 0.05$)”,因此,原文结论显然不妥。原作者采用重复测量数据方差分析进行组内比较,采用单因素方差分析进行组间比较,采用单因素设计资料方差分析对资料进行处理也是不妥的,同时也没有考虑温度与持续时间的交互作用。

本研究如果采用两因素析因设计的方差分析等多因素分析方法较为妥当。采用析因设计资料的方差分析,不仅能分析各因素的单独效应和主效应,还能分析各因素交互作用的效应,能够更准确地估计和控制误差的大小,因而结论更可靠。

【案例释疑】 试验设计按试验因素(干预因素)多少可分为单试验因素设计和多处理因素设计。单试验因素设计只有一个试验因素,通常采用完全随机设计、随机区组设计和拉丁方设计。多试验因素设计一般安排两个或两个以上试验因素,如析因设计、交叉设计等,此类设计试验因素间的交互作用往往是分析的重要内容,而单试验因素设计则不能分析交互作用,这是两种设计类型的不同之处。

单独效应是指其他因素的水平固定时,同一因素不同水平间的差别。主效应是指某一因素各水平间的平均差别。当某因素的各个单独效应随另一因素水平的变化而变化,则称这两个因素间存在交互效应或交互作用。

当某项试验研究中涉及两个或多个试验因素时,一般它们之间都存在一定程度的交互作用,此时,若采取“视而不见”的态度,很可能会得出错误的结论。正确的做法是通过预试验或借助他人的经验,对可能存在的交互作用在试验设计时就予以考虑。可考察交互作用的试验设计类型比较多,其中最为常用的是析因设计和正交设计。析因设计也叫全因子试验设计,它要求将全部试验因素的水平进行全面组合,每种组合称为一个试验点,在各试验点上要求至少做2次以上独立的重复试验。此设计需要的试验次数较多,但可以对各因素之间的各级交互作用的效应估计得比较准确。对于某些试验研究问题,若需要考察的试验因素较多,但需要考察的交互作用项较少时,按析因设计安排试验次数太多,不仅研究者无法承受,同时也有点浪费,此时,可考虑选用正交试验设计。在试验因素的水平数相同或呈二倍关系时,用正交设计取代析因设计,往往可以较大幅度地减少试验点数,而获得的试验结果又比较可靠,数据处理也比较方便。

3.3.4 误用单因素方差分析处理三因素资料

【案例3-7】 探讨不同牵伸载荷条件对体外人肌腱细胞纤维型肌动蛋白(F-actin)的影响,分析不同载荷条件与人肌腱细胞F-actin的变化关系。取5~7代人肌腱细胞加载不同牵伸载荷,其中牵伸强度分别为4%、8%、12%;时间为2、4、8、12、24h;频率为0.5Hz、1.0Hz。观察F-actin的变化、细胞核固缩及凋亡小体形成;图像分析软件测定单个细胞平均荧光强度。统计学处理:试验

所得各组细胞荧光强度以 $\bar{x} \pm s$ 表示, 采用 SPSS 13.0 统计软件, 进行单因素方差分析 (one-way ANOVA), 方差齐性检验后组间两两比较采用 LSD 检验, $P < 0.05$ 为差异具有统计学意义。

【案例辨析】 本试验中考虑了 3 个因素, 即强度、时间和频率, 皆为试验因素。一般来说, 它们之间可能存在一定程度的交互作用。采用单因素设计得出的所谓“最佳水平”, 放在一起试验时, 情况可能发生很大变化。当因素之间的交互作用很复杂且具有统计学意义时, 得出的结果可能很糟糕, 结果并不是各个单因素设计所得出的水平的简单组合。所以本文得出的结论是不可靠的。

试验中涉及 3 个试验因素, 一般可通过采用析因设计或正交设计来处理。本试验各因素所取的水平不一致, 没有相应的正交表, 可通过预试验的方式, 确定适当的水平段及其具体取值, 采用适宜的正交表进行试验。结果, 除了分析各因素的单独效应, 还分析出各因素之间部分交互作用的效应。

【案例释疑】 所谓正交设计, 是指利用一系列规格化的正交表来安排多因素试验的一种十分有效的设计方法, 可以分析各因素的主效应和低阶交互作用的效应。正交设计是通过进行表头设计来实现的, 通常先根据需考察的因素数、水平数及交互作用的个数 (以专业知识为依据), 选择合适的正交表, 每个因素的效应对应于正交表的 1 列, 各因素的交互作用所在的列通过查找所对应的交互作用表来获得, 它对应于正交表的一列或多列上。此外, 还需要空出若干列来, 在方差分析时提供误差项用于统计分析。

3.3.5 非参数检验与 Fisher 精确检验

【案例 3-8】 比较快速人工心脏起搏与硝普钠在胸主动脉腔内修复治疗中的安全性、有效性及对移植物流定位准确性的影响。研究入选 2007 年 9 月至 2009 年 2 月在广东省人民医院心内科因胸主动脉疾病择期行胸主动脉腔内修复治疗 (TEVAR) 的 197 例患者 (夹层 175 例, 动脉瘤 22 例)。随机分为硝普钠组 ($n = 98$) 和快速人工心脏起搏组 ($n = 99$), 在腔内修复治疗术中分别应用上述两种方法以助于移植物的准确定位与释放。比较两组的血流动力学指标、移植物流定位及释放准确性、手术时间、手术前后肾功能和神经认知功能变化以及内漏和截/偏瘫发生率。结果快速人工心脏起搏组所有患者均成功植入右室起搏电极, 并成功实施快速人工心脏起搏。快速人工心脏起搏组的主动脉压 (mmHg, $1 \text{ mmHg} = 0.133 \text{ kPa}$) 明显低于硝普钠组 [(47 ± 5) 比 $(82 \pm 7) \text{ mmHg}$, $P = 0.003$], 其血压恢复时间 [$(9 \pm 2) \text{ s}$ 比 $(481 \pm 107) \text{ s}$, $P < 0.01$] 及手术时间 [$(94 \pm 16) \text{ min}$ 比 $(103 \pm 24) \text{ min}$, $P < 0.01$] 均明显短于硝普钠组, 并且快速起搏组移植物流定位及释放准确性明显高于硝普钠组 [$(2 \pm 1) \text{ mm}$ 比 $(4 \pm 3) \text{ mm}$, $P < 0.01$]。两组患者术前、术后的肾功能及神经认知功能和内漏、截/偏瘫发生率差异无统计学意义。结论与硝普钠相比, 快速人工心脏起搏可安全应用于 TEVAR, 缩短手术时间, 还有助于移植物流的准确定位与释放。

统计学方法: 所有计量资料以 $\bar{x} \pm s$ 表示。计数资料之间的比较用 χ^2 检验, 计量资料之间的比较用 t 检验, 肾小球滤过率 (eGFR) 及 MMS 评分不符合正态分布则应用秩和检验进行统计学分析。以 $P < 0.05$ 为差异有统计学意义。

结果: 两组内漏、截/偏瘫及围手术期死亡发生率比较: 硝普钠组与快速起搏组比较, 内漏为 8 例 (占 8.2%) 比 6 例 (占 6.1%); 截/偏瘫为 6 例 (占 6.1%) 比 4 例 (占 4.0%); 围手术期死亡为 2 例 (占 2.0%) 比 1 例 (占 1.0%); 差异均无统计学意义 (均 $P > 0.05$)。

【案例辨析】 原作者采用“ t 检验”方法比较计量资料的组间差别合适吗? 作为参数检验方法, t 检验和方差分析有其应用前提条件, 只有满足独立性、正态性和方差齐性检验条件后才能应用。“肾小球滤过率 (eGFR) 及 MMS 评分由于不符合正态分布和方差齐性的要求, 则应用秩和检验进行

统计学分析”。从文中可见,低血压持续时间、血压恢复时间等也不符合正态分布、总体方差相等的假设,因此,对该指标的分析同样不能应用 t 检验,而应该采用秩和检验方法分析。

针对定性资料的比较,原作者采用卡方检验分析。本资料为四格表资料,只有当样本总数 $n > 40$ 且无小于 5 的理论频数时,才适合用一般卡方检验。本文结果显示围手术期死亡发生率等很低,有小于 5 的理论频数,因此不适合用一般卡方检验,不能直接采用卡方检验方法进行比较。对于 $n < 40$ 或理论频数 < 1 的四格表资料,应选用 Fisher 精确检验法。正确的统计分析方法应该是采用校正卡方检验或者 Fisher's 精确检验方法来比较组间差别。

【案例释疑】 若资料不满足参数检验的前提条件,可选用适合分析偏态分布资料的非参数检验(如秩和检验)进行分析,或找到满足前述两个前提条件的相应变量变换方法后,再用相应的参数检验法。方差齐性检验主要用于:①两组或多组变量值间变异度的比较;②两个或多个总体均数间比较时,须先进行方差齐性检验,若方差齐,可用 t 检验或方差分析,否则可用变量变换的方法,使之方差齐后再用 t 检验或方差分析,或用对方差没有特别要求的 Z 检验、 F 检验及其他非参数的统计分析方法。

常规的统计推断方法,如总体均数的区间估计、成组设计的两总体均数比较等,通常都要求样本所来自的总体分布为已知的函数形式,在此基础上,对其中的未知参数进行估计或检验,称为参数统计。但在许多实际问题中,总体分布的函数形式往往不知道,或者知道得很少,如只知道总体分布是连续型的或离散型的。这时参数统计分析方法就不大适用了,而需要借助于另一种不依赖总体分布的具体形式,也不对参数进行估计或检验的统计分析方法,称为非参数统计,其检验方法就是非参数检验。归纳起来,非参数检验适用于下列情况。

(1)等级资料,如用某药治疗老年慢性气管炎,其疗效分为控制、显效、有效、无效四个等级。

(2)偏态分布资料。当观察资料呈偏态分布,而又未经变量变换或虽经变换但仍未达到正态或近似正态分布时,宜选用非参数检验分析。

(3)未知分布类型的资料。生物医学领域尚有不少观察资料分布类型未知,可用非参数检验处理。

(4)要比较的各组资料变异度相差较大,方差不齐,且不易通过变换达到齐性,宜用非参数检验分析。

(5)初步分析有些医学资料由于统计工作量过大,可采用非参数统计法进行初步分析,挑选其中有意义者再做进一步分析。

(6)对于一些特殊情况,如从几个总体所获得的数据,往往难以对其原有总体分布作出估计,在这种情况下,宜采用非参数统计法。

(7)资料中含有不确定数据的情况,如测量结果只知道大于某一数值,但具体数值不知道,这种情况下,参数统计方法无法分析,宜采用非参数统计法。

非参数检验的首要优点是对资料的要求不像参数检验那样严,能适于处理非正态、方差不齐,以及分布型不明确等类型的资料。其次,有些问题本身没有适当的参数检验方法可用,而非参数检验法则能予以处理。

非参数检验的不足之处是,由于它对原始数据中所包含的信息利用得不充分,检验效能降低。尤其是对适宜用参数检验法处理的资料,若用非参数检验法分析,常损失部分信息,降低检验效能,一般犯第二类错误的概率 β 比参数检验大。故适宜用参数检验法的资料,应首选参数检验法。需要指出的是,若参数检验的应用条件得不到满足,则用非参数检验才是明智的选择。

定性资料的卡方检验方法注意事项:

分类变量资料中,随机变量 X 呈非连续型分布,而卡方分布属于连续型分布,当 $\nu = 1, 1 < T < 5$

时,非连续型随机变量 X 的分布与卡方分布偏离较大,于是英国统计学家 Yates 提出了简单的校正项,即连续性校正。

在四格表资料卡方检验中,若有一格 $T < 1$, $n < 40$,尤其用其他检验方法所得概率接近检验水准时,宜用四格表精确概率法即 Fisher 精确检验进行分析。本法的基本思想是在所收集数据周边合计保持不变的条件下,求出 n 例中有 X 例发生阳性时,四个基本数据各种组合中的极端情况累积概率。

3.3.6 相关与回归分析误用

【案例 3-9】 分析参加健康体检人群中视网膜动脉硬化的患病率和相关因素。计算 2010 年 1 至 10 月在北京协和医院健康体检人群中视网膜动脉硬化的患病率,并通过多因素 Logistic 回归模型分析与该病相关的危险因素。结果研究期间共有非糖尿病成年人群 17 886 名进行眼科检查并纳入分析,诊断视网膜动脉硬化 1721 例,总患病率为 9.6%,以 1 级(88%)和 2 级(11%)病变为主。视网膜动脉硬化的患病率随年龄增长和人群中心血管病危险因素(超重/肥胖、高血压、高脂血症、空腹血糖升高)的个数增加而增高。在多因素 logistic 回归分析中,年龄增大、男性、超重或肥胖、高血压和高脂血症均为视网膜动脉硬化的独立危险因素。

【案例辨析】 原作者采用 logistic 回归方法分析危险因素,但是统计分析方法中并未提及是如何构建多重 logistic 回归模型的,未交代筛选自变量的方法。

文中提到“视网膜动脉硬化的患病率随年龄增长和人群中心血管病危险因素(超重/肥胖、高血压、高脂血症、空腹血糖升高)的个数增加而增高”,但是并没有进行相关分析,因此,该结论并无统计分析支持。如果单纯统计分析比较组间有无差别,那么结论仅限为差异有无统计学意义,而不能推断它们之间的关系。建议进行相关分析后,再下此结论。

【案例释疑】 在双变量关系分析中,有时研究者所关心的是分析两变量间是否呈线性关系?若线性关系存在,其相关的方向和密切程度如何?通常将分析两变量间是否呈直线关系的方法称为直线相关分析,也称为简单相关分析。如果分析两变量间呈直线趋势的依存关系,则称为直线回归分析。至于两变量间究竟呈直线关系、曲线关系还是无任何确定性的关系,可用散点图来直观判断。

应用相关与回归分析时应注意的问题包括以下几个方面。

(1) 首先应考察受试对象是否具有同质性,即就所观测的两个定量指标而言,受试对象是否来自单组设计总体的一个随机样本。

(2) 做相关与回归分析时要有实际意义,不能对毫无关联的两个变量作相关、回归分析,必须对两个变量的内在联系有所认识,即应有专业依据。

(3) 在进行相关与回归分析之前,应先绘制并正确分析散点图。当观察点的分布有直线趋势时,才适宜作直线相关和/或直线回归分析;如散点图呈明显的非直线趋势,应使之直线化再进行分析或直接拟合相应的曲线回归方程;若散点图显示全部散点呈一个圆盘状分布,表明两变量之间没有任何确定性关系;散点图还能提示资料中有无异常点。

(4) 直线回归方程的适用范围一般以自变量的取值范围为限,若无充分理由证明超过自变量取值范围外还是直线,应避免外延。

(5) 经统计学检验得出直线相关或/和直线回归分析有统计学意义时,还需考察两变量之间的关系是否具有实际意义,当 $Pr^2 > 0.5$ 时才认为直线相关或/和直线回归分析不仅有统计学意义,而且也具有一定的实际价值。

(6) 相关关系不一定是因果关系,也可能是伴随关系,有相关关系不能证明事物间确有内在联系。是否有因果联系应通过试验验证。

除了直线相关关系外,变量之间也可能呈某种曲线关系。除了双变量分析,还有多变量分析(通常为一个反应变量如何随多个自变量变化而变化的规律的研究,称为多重回归分析)。实际应用中,需要根据实际情况选择合适的回归分析方法。

3.4 统计结果报告中的常见问题及案例辨析

3.4.1 假设检验的结果表达

【案例3-10】 分析麻痹性痴呆患者边缘系统相关结构的MRI表现,进一步探讨该病的病理机制,同时提出对该病MR诊断的思路。回顾性分析2006年5月至2010年6月中山大学第三附属医院收治的31例麻痹性痴呆患者,对其边缘系统相关结构的MR影像进行分析,包括体积和信号的改变。统计学处理:应用SAS 8.1统计软件,采用 t 检验对数据进行统计学分析,比较病例组与对照组的两侧颞角面积的差异, $P < 0.05$ 为差异有统计学意义。结果:31例患者边缘系统相关结构的异常表现。左侧颞角面积:病例组为 $(193 \pm 181) \text{ mm}^2$,正常组为 $(43 \pm 23) \text{ mm}^2$ 。右侧颞角面积:病例组为 $(172 \pm 115) \text{ mm}^2$,正常组为 $(38 \pm 17) \text{ mm}^2$ 。两组数据 t 检验均 $P < 0.05$ 。

【案例辨析】 本研究病例组与正常组的比较,由于不满足 t 检验的应用条件(各总体呈正态分布与各总体方差相等),应此不能采用 t 检验,而应该采用非参数检验。除此之外,在统计分析结果报告中,如果遇到 $P < 0.05$,不能仅仅写 $P < 0.05$ 。正确的做法应交代资料的设计类型,根据分析目的和资料所具备的前提条件选择合适的统计分析方法,并给出统计量的数值和精确的 P 值,以增加论文的科学性和结论的可靠性。

【案例释疑】 统计学差异的假设检验方法很多,其基本步骤一般是:假设对比各组的总体参数相等或分布相同,无实质性差异,称为原假设,再根据原始数据计算出符合特定分布规律的统计量的大小。由统计量大小再推出所对应的 P 值。 P 值的意义是因抽样误差而出现此种程度差异的概率:若 P 值甚小(如 $P \leq 0.05$),则根据“小概率事件在—批试验(获得检验统计量的一个具体取值)中不可能发生”这一原理,拒绝原假设,从而认为差异具有统计学意义;否则,若 P 值不算小,就不拒绝原假设,认为差异不具有统计学意义。应当说明的是:这种差异采用有无统计学意义的说法来描述,而不宜理解成实际的差距大小,即不适合用差别显著或非常显著的说法来描述。这是因为统计量的计算一方面取决于两样本的差异,另一方面还与它们的抽样误差有关,抽样误差很小(如样本量很大),即使这种差异在专业上并无实际意义,也有可能得出差异在统计学上有意义的结果。反之,若抽样误差较大(如样本量较小,检验效能低),即使在总体中实际上存在差异,统计学上并不一定能发现这种差异。

研究者在结果解释时不应仅以 P 值而妄下定论,作结论也不宜绝对化。“ $P = 0.049$ ”与“ $P = 0.051$ ”没有本质的差别。因 P 值与样本含量、数据的变异度等有关,应根据犯第一类错误的概率、检验效能及专业知识,合理确定样本量,以便能获得较为可靠的推断。鉴于计算机及统计软件的普及,研究者在论文中应给出统计量及 P 的具体数值,以便让读者对结果的可靠性做出自己的评价。

因此,假设检验的结果表达,不能仅仅给出 P 值范围,还要求给出检验统计量的实际值(如 Z 值、 t 值、卡方值等)和具体的 P 值。描述统计量,如均数、率、相关系数,无论检验结果是否有统计学意义,均应列出,并且指明对哪些指标所对应的参数已进行过统计学检验。若用符号(如用“NS”表示 $P > 0.05$,”*”表示 $P < 0.05$,”**”表示 $P < 0.01$)表示 P 值大小,要加以说明和统一。检验水准 α 值传统上习惯取0.05和0.01两个界值,现在随着计算机和统计软件的普及,提倡在检验结果的表达时给出具体的 P 值,如 $P = 0.012$ 或 $P = 0.361$ 等,这样可以为读者提供更充分和精确的信息,对研究结论的统计学证据认识更为详尽,同时也可给其他同类研究提供更加令人可信和再分析的依据,例如用于Meta分析。

3.4.2 统计表使用的错误辨析与释疑

【案例 3-11】 比较 4 种皮质下单灶性卒中患者的认知损害的特点，了解丘脑和基底节的认知功能，为早期识别血管性认知功能损害(VCI)患者提供依据。结果列于表 3-2，该表格是否符合统计规范？应如何修改？

表 3-2 皮质下单灶卒中组与健康对照组一般情况 ($\bar{x} \pm s$)

组 别	例 数	年龄(岁)	受教育年限(年)	男:女	MMSE 总分	CESD 总分
左丘脑组	14	63 ± 8	6.9 ± 3.1	10:4	24.9 ± 2.9	8.3 ± 6.8
左底节组	17	59 ± 6	7.4 ± 2.0	12:5	26.5 ± 3.1	5.3 ± 6.8
右丘脑组	15	62 ± 8	7.5 ± 4.2	11:4	26.2 ± 3.4	9.3 ± 8.3
右底节组	17	60 ± 6	6.9 ± 3.6	12:4	25.6 ± 2.9	10.0 ± 8.6
对照组	34	61 ± 8	8.1 ± 3.1	27:7	28.1 ± 1.7	7.9 ± 7.6
F 值(P 值)		0.642(0.643)	0.543(0.705)	0.215(0.929)	4.785(0.002)	0.913(0.460)

【案例辨析】 本文的统计表存在一个不足之处，在表格的最底行出现统计量 F 值与 P 值(应当指出的是性别构成比较时所用统计量并非 F 统计量)，这样的安排并不合理，应删除最后一行，统计分析结果可以列于统计表下方备注中或者在论文中体现。修改后的统计表略。

【案例释疑】 统计表的结构：(1)标题，位于统计表的正上方，简明扼要地将统计表的主要内容表达出来。(2)标目，用以说明表格内的项目。根据其位置与作用可分为横标目、纵标目和总标目。横标目位于统计表的左侧，用于说明各行数据，可被称为“主语”；纵标目说明各纵栏数字的含义，可被称为“谓语”。(3)线条，统计表内的线条不宜过多，常常仅包括三条基本线，即顶线、底线、纵标目分割线。统计表左右两侧不应有边线。表的左上角不应有斜线，表内不应有竖线。(4)数字，表内的数字必须准确无误，用阿拉伯数字表示。同一指标的小数位数应一致。上下对齐，表内不留空格，数字暂缺或未记录用“…”表示，无数字用“-”表示。有相对数时，应同时给出绝对数，以便读者了解和核算。(5)备注，不是统计表固有的组成部分，一般不列入表内。如需对某个数字或指标加以说明，可在该数字或指标右上方用“*”之类的符号标注，并在统计表的下方用文字加以说明。

3.4.3 统计图使用的错误辨析与释疑

【案例 3-12】 在《常规体检人群中视网膜动脉硬化的患病率和相关因素分析》一文中，原作者分析了视网膜动脉硬化的患病率和相关因素，得出视网膜动脉硬化的患病率随年龄的增加而增高(图 3-1)。该统计图是否符合统计规范？应如何修改？

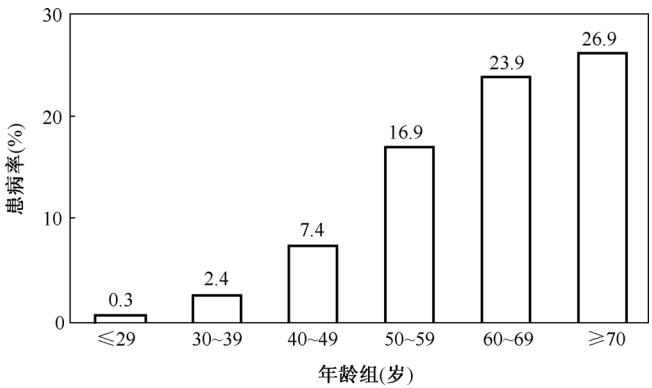


图 3-1 不同年龄组的体检人群中视网膜动脉硬化的患病率变化趋势

【案例辨析】 本文的统计图是条图或称为柱状图，一般用来直观表达分类资料各组之间指标的数量大小，而不反映事物的动态变化。若用线图(见图 3-2，用图域表示更好)来表达视网膜动脉硬化的患病率随年龄的增加而增高，更为恰当。统计图可以给人直观深刻的印象，较统计表表达资料显得生动活泼。

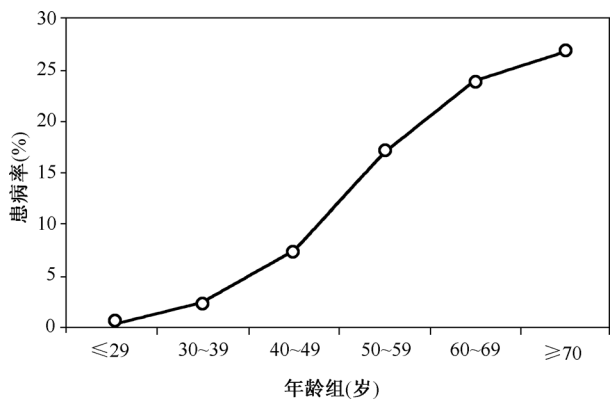


图 3-2 不同年龄组的体检人群中视网膜动脉硬化的患病率变化趋势

【案例释疑】 常用统计图的名称、适合的变量类型及用途见表 3-3。

表 3-3 常用统计图的用途

统计图形	变量类型	用途
条图	连续型变量，分类变量	用直条的长短表达数值大小
百分条图	分类变量	用长条各段的长度(面积)表达所占比例
圆图	分类变量	用圆中扇形面积表达所占比例
线图	连续型变量，分类变量	用线段的升降表达事物的动态变化(绝对差值)
半对数线图	连续型变量，分类变量	用线段的升降表达事物的发展速度(相对比)
直方图	连续型变量	用直条的高度或面积表达各组段的频率或频数
散点图	双变量，连续型变量	用点的排列趋势和密集度表示两变量的相关关系
箱式图	连续型变量	用“箱”和“触须”等的位置表示变量的分布特征

3.5 本章内容小结

在本章中，结合文献具体实例，从统计学角度，简单介绍了如何在医学研究中正确运用统计思想、理论与方法，提高识别“科研设计、统计表达与描述、统计分析”三方面出现错误的能力及其正确实施的理论指导和技术支持。对于《中华医学杂志》的撰稿，统计学处理的有关要求如下。

(1)统计研究设计：应交代统计研究设计的名称和主要做法。如调查设计(分为前瞻性、回顾性或横断面调查研究)；试验设计(应交代具体的设计类型，如自身配对设计、成组设计、交叉设计、析因设计、重复测量设计、正交设计等)；临床试验设计(应交代属于第几期临床试验，采用了何种盲法措施等)。主要做法应围绕四个基本原则(随机、对照、重复、均衡)概要说明，尤其要交代如何控制重要非试验因素的干扰和影响，还应特别重视和认真做好试验或调查过程中的质量控制。

(2)资料的表达与描述：用 $\bar{x} \pm s$ 表达近似服从正态分布(至少也应该是单峰对称分布)的定量资料，用 $M(Q_1 \sim Q_3)$ 表达呈偏态分布的定量资料；用统计表时，要合理安排纵横标目，并将数据的含义表达清楚；用统计图时，所用统计图的类型应与资料性质相匹配，并使数轴上刻度值的标法符合数学原则；用相对数时，分母不宜小于 20，要注意区分百分率与百分比。

(3) 统计分析方法的合理选择: 对于定量资料, 应根据所采用的设计类型、资料所具备的条件和分析目的, 选用合适的统计分析方法, 不应盲目套用 t 检验和单因素方差分析; 对于定性资料, 应根据所采用的设计类型、定性变量的性质和频数所具备的条件以及分析目的, 选用合适的统计分析方法, 不应盲目套用卡方检验。对于回归分析, 应结合专业知识和散布图, 选用合适的回归类型, 不应盲目套用简单直线回归分析, 对具有重复试验数据的回归分析资料, 不应简单化处理; 对于多因素、多指标资料, 要在一元分析的基础上, 尽可能运用多元统计分析方法, 以便对因素之间的交互作用和多指标之间的内在联系进行全面、合理的解释和评价。

(4) 统计结果的解释和表达: 当 $P < 0.05$ (或 $P < 0.01$) 时, 应说明对比组之间的差异有统计学意义, 而不应说对比组之间具有显著性 (或非常显著性) 的差别; 应写明所用统计分析方法的具体名称 (如成组设计一元定量资料的 t 检验、两因素析因设计一元定量资料的方差分析、多个均数之间两两比较的 q 检验等), 应给出统计量的具体值 (如 $t = 3.45$, $\chi^2 = 4.68$, $F = 6.79$ 等), 还应尽可能给出具体的 P 值 (如 $P = 0.0238$); 当涉及总体参数 (如总体均数、总体率等) 时, 在给出假设检验结果的同时, 再给出 95% 置信区间。

参 考 文 献

- [1] 胡良平, 李子建. 医学统计学基础与典型错误辨析. 北京: 军事医学科学出版社, 2003.
- [2] 孙振球. 医学统计学 (第 2 版). 北京: 人民卫生出版社, 2006.
- [3] 方积乾. 卫生统计学 (第 6 版). 北京: 人民卫生出版社, 2008.
- [4] 中华医学杂志稿约. 中华医学杂志, 2011, 91(25): 1799-1800.
- [5] 本刊对来稿中统计学处理的有关要求. 中华医学杂志, 2011, 91(27): 1894.
- [6] 黄文晖, 何鹏程, 罗建方等. 快速起搏用于胸主动脉覆膜支架精确释放的临床随机对照研究, 中华医学杂志, 2011, 91(24): 1668-1672.
- [7] 刘鑫, 刘启峰, 徐晔等. 中国年轻乳腺癌的临床病理特征及预后分析. 中华医学杂志, 2011, 91(26): 1817-1820.
- [8] 吴莉娅, 秦超, 叶子明等. 64 层螺旋 CT 对椎动脉狭窄诊断价值的荟萃分析. 中华医学杂志, 2011, 91(25): 1766-1769.
- [9] 严强, 倪俊, 张国雷等. 过渡肠内营养对家兔重症急性胰腺炎血液流变学及炎症因子的影响, 中华医学杂志, 2011, 91(28): 2006-2010.
- [10] 段敏超, 钟小宁, 黄颖等. 香烟烟雾暴露肺气肿小鼠肺中 Th17 细胞的变化及可能机制. 中华医杂志, 2011, 91(28): 1996-2000.
- [11] 印红梅, 李安宝, 余健. 不同温度与持续时间下射频热凝对兔半月神经节的组织学影响. 中华医学杂志, 2011, 91(24): 1718-1721.
- [12] 邓银栓, 唐康来, 谢关明等. 不同牵伸载荷对体外培养人肌腱细胞纤维型肌动蛋白的影响. 中华医学杂志, 2011, 91(25): 1780-1785.
- [13] 吴婵董, 方田, 楼慧萍等. 常规体检人群中视网膜动脉硬化患病率和相关因素分析. 中华医学杂志, 2011, 91(26): 1821-1824.
- [14] 江婷, 陈少琼, 单鸿等. 麻痹性痴呆边缘系统相关结构损害的 MRI 表现. 中华医学杂志, 2011, 91(29): 2055-2059.
- [15] 龚文萃, 丁关萍, 郭起浩等. 四种不同脑部位卒中后认知功能状况比较. 中华医学杂志, 2011, 91(27): 1904-1908.

第4章 医学遗传学统计设计与数据分析

《中华医学遗传学杂志》主要收录我国医学遗传学和人类遗传学以及相关领域中的基础理论、临床实践、技术方法等的最新研究成果。

本章针对该杂志近5年论文中科研设计与统计分析出现的较严重问题进行辨析与释疑,包含了研究设计、定量资料统计分析、定性资料统计分析、简单相关回归和多因素分析等方法。其中该杂志统计问题以定性资料统计分析及其多重比较、非条件 logistic 回归分析等最为多见,本章也将对该类方法出现的问题重点阐述,为广大医学遗传学科研工作者提供参考。

4.1 科研设计错误辨析

4.1.1 研究对象选择不当

【案例4-1】 原文目的是探讨白细胞介素1B基因启动子区-511位点C/T基因多态性与冠心病严重程度的相关性。原作者选择某医院127例冠心病(CHD)患者,其中急性冠脉综合征(ACS)78例,稳定劳累性心绞痛(SA)49例;选择对照组152人。所有研究对象均为汉族自然人群,其无血缘关系。检测所有研究对象的IL1B-511位点C/T基因型,采用酶法测定血脂各项水平。检测分析的结果为IL1B-511位点C/T基因多态性在急性冠脉综合征(ACS)组和对照组间分布差异存在统计学意义($P < 0.01$),CT及TT基因型患者患ACS的相对风险度约是CC基因型的2.56倍(95%可信区间=1.17~5.59)。由此获得结论是IL1B-511位点C/T基因多态性与CHD的严重程度存在相关。

【案例辨析】 原文探讨IL1B-511位点C/T基因多态性与冠心病严重程度的相关性,研究对象应该只限定于冠心病人群,再将冠心病人群按严重程度分成若干组,而原文选取了冠心病患者和对照组,且对照组是否为“健康人群”也没做任何介绍,只是依据临床表现将冠心病(CHD)患者分成急性冠脉综合征(ACS)组和稳定劳累性心绞痛(SA)组。最后根据IL1B-511位点C/T基因多态性在急性冠脉综合征(ACS)组和对照组间分布差异存在统计学意义就得出IL1B-511位点C/T基因多态性与CHD的严重程度存在相关的结论很难令人信服。IL1B-511位点基因型在ACS组和对照组间分布差异有统计学意义只能说明IL1B-511位点基因多态性和冠心病存在关联,并不能说明它和冠心病严重程度的相关性。所以研究对象一定要反映研究目的。另外,就“IL1B-511位点C/T基因多态性与CHD的严重程度”资料而言,一个为多值名义变量、另一个为多值有序变量,很可能采用的是四格表或多行多列表的差异性分析方法处理了资料,怎么又得出了“相关性”结论?

【案例释疑】 探讨某位点基因多态性与冠心病严重程度的关系,应该按照疾病的严重程度分层后,再从各层中随机抽取足够数量的冠心病患者作为研究对象。未采用相关分析方法处理资料,不应下存在相关性的结论。分析方法与结论应具有相合性。

【案例4-2】 原文目的是对中国人CHRNA4基因的多态性分析。方法是随机抽取100名北京地区流行病学调查老年人和100例原发性帕金森病患者,测序确定CHRNA4基因各多态位点基因频率。

【案例辨析】 原文研究目的是研究中国人 CHRNA4 基因的多态性，但是研究对象选取的是原发性帕金森病患者作为病例组，流行病学调查的老年人作为对照组。很明显研究对象的选取并不反映研究目的。

【案例释疑】 应根据研究目的，确定研究对象的总体，并制订合理的纳入、排除和剔除标准。在特定的总体中，采用合适的随机抽样方法抽取足够数量的个体组成研究样本。在不同地区抽取样本时，应特别强调样本的代表性。应弄清楚：是希望做中国人 CHRNA4 基因的多态性的本底调查，还是希望研究几类不同人群 CHRNA4 基因的多态性的频数分布规律及其他他们之间的差异。研究目的不同，考察的影响因素不同，对照的形式可能不一样，拟采取的调查设计类型有别，对资料的统计分析方法也不尽相同。

4.1.2 样本含量太少

【案例 4-3】 原文是关于 Wnt 通路相关基因多态性与胃癌发病风险的病例对照研究。选取来自某医院 26 例胃癌患者作为病例组，其中男 15 例，女 11 例；33 例慢性浅表性胃炎患者作为对照组，其中男 22 例，女 11 例。检测所有研究对象的 rs3755557 位点及 rs1880481 位点基因型。rs3755557 位点基因型及等位基因频率在病例组和对照组中的分布见表 4-1。

表 4-1 GSK-3 β 基因 rs3755557 多态基因型及等位基因频率在病例组和对照组中的分布

		样本量(26)	男(15)	女(11)	样本量(33)	男(22)	女(11)
		胃 癌 组			对 照 组		
基因型 频率(%)	AT	5(19.33)	2(13.33)	3(27.27)	7(21.21)	5(22.73)	2(18.18)
	AA	20(76.92)	12(80.00)	8(72.73)	26(78.79)	17(77.27)	9(81.82)
	TT	1(3.85)	1(6.67)	0	0	0	0
	AA + AT	25(96.15)	14(93.33)	11(100.00)	33(100.00)	22(100.00)	11(100.00)
	AT + TT	6(23.08)	3(20.00)	3(27.27)	7(21.21)	5(22.73)	2(18.18)
	AA + TT	21(80.77)	13(86.67)	8(72.73)	26(78.79)	17(77.27)	9(81.82)
等位基因 频率(%)	A	45(86.54)	26(86.67)	19(86.36)	59(89.39)	39(88.64)	20(90.91)
	T	7(13.46)	4(13.33)	3(13.64)	7(10.61)	5(22.73)	2(9.09)

【案例辨析】 原文目的是研究 Wnt 通路相关基因多态性和胃癌发病风险的关系，但研究对象病例组选取了 26 例，对照组选取了 33 例，再按基因型和性别分组，有些组样本含量太少，代表性很差，结论的可信度也就无从谈起了。

【案例释疑】 根据病例对照研究样本含量估计的方法计算样本量。

4.1.3 均衡原则方面存在的问题

【案例 4-4】 原文目的探讨广东地区汉族人人类白细胞抗原 - DQB1 等位基因与十二指肠溃疡的遗传关联。方法是选取某医院十二指肠溃疡(DU)患者 105 例，其中男 72 例，女 33 例，均为广东地区汉族人；对照组选取 105 例健康人，年龄相当、无消化性溃疡及其他特殊病史或家族史，其中男 51，女 54 例。组间等位基因分布差异用卡方检验或 Yate’s 校正法及 Fisher 确切概率检验。结果显示 HLA-DQB1 等位基因分布在十二指肠溃疡患者和正常人之间存在差异。

【案例辨析】 从原文可以看出在十二指肠溃疡患者和正常对照之间性别分布不均衡，可能存在较大差异，而且文中也没有检验两组人群之间性别分布是否存在差异，所以性别分布的不均衡可能会影响 HLA-DQB1 等位基因和十二指肠溃疡关联结论的可靠性。

【案例释疑】 在选取研究对象时应该遵循均衡的原则，可能的影响因素在比较的两组之间应

该尽可能均衡可比,且需要经过统计学检验其差异没有统计学意义,如差异有统计学意义应该采用分层分析或多因素分析方法控制混杂。

【案例4-5】 原文目的研究脂蛋白脂酶 Ser447Ter 基因多态性与动脉粥样硬化性脑梗死(CI)发病的关系及其对血脂水平、颈动脉斑块的影响。方法是选取某医院神经内科动脉粥样硬化性脑梗死住院患者,共166例,其中男113例,女53例;对照组为该院门诊体检健康成人,共72例,其中男32例,女40例。CI组和对照组在年龄、性别、体重指数、腰围/臀围比的差异无统计学意义。计量资料用 t 检验、方差分析,计数资料用 χ^2 检验。结果显示脂蛋白脂酶 Ser447Ter 基因多态性与血脂变化及脑梗死的关系密切,G等位基因可能是脑梗死的保护基因型。

【案例辨析】 从原文可以看出病例组和对照组间性别分布不均衡,但原文结果显示CI组和对照组在性别上的差异无统计学意义。

【案例释疑】 根据原文数据做病例组和对照组间性别比较,经 χ^2 检验得卡方值为11.78($P < 0.05$),病例组和对照组间性别分布差异有统计学意义。所以在进行Ser447Ter基因多态性与动脉粥样硬化性脑梗死关系研究时应按性别分层或进行多因素分析。

4.2 定量资料统计分析中存在的问题

4.2.1 误用 t 检验处理单因素多水平设计定量资料

【案例4-6】 原文研究目的探讨PAI-1启动子区4G/5G多态性与IgA肾病的发生、进展和临床表现的关系。研究对象为某院IgA肾病296例,正常健康对照310例,收集IgA肾病患者的临床资料,检测所有研究对象的PAI-1基因4G/5G多态性。结果显示PAI-1基因型分布在IgA肾病组和正常对照组间差异无统计学意义。该研究又将IgA肾病患者按照PAI-1基因4G/5G各基因型分成三组:分别为4G4G组、5G5G组和4G5G组,对这三组间临床特征进行了比较,比较指标见表4-2。统计方法叙述为用SPSS软件对临床资料进行分析,计量资料用 t 检验,计数资料用 χ^2 检验。结果显示4G4G组IgA肾病患者的肌酐清除率明显低于4G5G组;4G4G组患者的血清总胆固醇明显高于5G5G组。

表4-2 4G/5G各基因型间临床特征的比较

项 目	4G4G 组	5G5G 组	4G5G 组
例数	98	56	142
年龄(岁)	32.4 ± 11.2	30.3 ± 11.9	31.4 ± 11.9
性别(男/女)	63/35	40/16	92/50
肌酐清除率(mL/min)	66.4 ± 33.1 *	74.5 ± 42.9	79.7 ± 40.7
血清肌酐(μmol/L)	132.9 ± 117.9	115 ± 76.3	111.9 ± 63.4
血清白蛋白(g/L)	37.0 ± 5.2	37.8 ± 6.1	38.2 ± 6.8
血清总胆固醇(mmol/L)	2.08 ± 1.52 **	1.66 ± 0.91	1.85 ± 1.44
血清甘油酯(mmol/L)	5.34 ± 1.59	5.39 ± 1.82	5.37 ± 2.48
高血压(%)	41.8	39.3	35.9
大量蛋白尿(%)	28.9	23.2	22.1
低白蛋白血症(%)	27.6	23.2	20.4
血清高总胆固醇(%)	53.1 *	37.5	33.3
血清高三酰甘油(%)	30.6	35.7	33.3

注: * $P < 0.05$; * * $P < 0.01$ 。

【案例辨析】 从原文可以看出比较的指标中年龄、肌酐清除率、血清肌酐、血清白蛋白、血清总胆固醇和血清甘油酯都属于定量资料，而且是4G4G组、5G5G组和4G5G组三组间的比较，而在统计描述部分定量资料的比较只交代了用 t 检验，但 t 检验方法只能用于两组间的比较，三组间比较并不适用 t 检验。

【案例释疑】 该资料三组间定量资料的比较属于单因素三水平设计定量资料统计分析，若满足参数检验的前提条件(即正态性、独立性和方差齐性)，则应选用单因素三水平设计定量资料方差分析，否则应选用此设计定量资料 Kruskal- Wallis H 检验。当三组间差别具有统计学意义时，才适合进行两两比较。两两比较时不宜直接采用 t 检验，因为那样做会增大犯 I 类错误的概率。常可选用的两两比较方法有 q 检验(参数检验)、Nemenyi 法检验(非参数检验)等。

【案例 4-7】 原文目的研究汉族人群的一氧化氮合酶(NOS)基因 NOS3 - 922A/G 和 NOS3 - 894G/T 以及 NOS2 - 1173C/T 3 个位点的单核苷酸多态性与静息心率的相关性。随机选取了自然健康个体 211 例。检测上述位点的单核苷酸多态性。统计分析方法分类资料用卡方检验，计量资料采用 t 检验， $P < 0.05$ 为差异有统计学意义。结果显示 NOS3 - 922A/G 的各基因型 AA、AG、GG 静息心率比较，发现携带 AA 基因型者静息心率较 GG 者高，差异有统计学意义($P < 0.01$)；NOS3 - 894G/T 各基因型静息心率比较，发现携带 GG 基因型者静息心率较 TT 者高，差异有统计学意义($P < 0.05$)；NOS2 - 1173C/T 各基因型携带者的静息心率比较，差异均无统计学意义($P > 0.05$)。具体数据见表 4-3。

表 4-3 NOS3-922A/G、NOS3-894G/T 与 NOS2-1173C/T SNP 各基因型频率及其静息心率

基 因 型		例 数	基因型频率(%)	静息心率($\bar{x} \pm s$, 次/分)
NOS3 - 922	AA	111	52.61	72.52 \pm 9.89
	AG	90	42.65	70.97 \pm 8.42
	GG	10	4.74	61.50 \pm 8.61
NOS3 - 894	GG	109	51.66	71.84 \pm 8.92
	GT	81	38.39	70.89 \pm 12.82
	TT	21	9.95	66.87 \pm 10.45
NOS2 - 1173	CC	129	61.14	71.37 \pm 9.70
	CT	67	31.75	71.59 \pm 9.69
	TT	15	7.11	72.15 \pm 7.39

【案例释疑】 该资料每个位点三种基因型携带者的静息心率比较属于单因素三水平设计定量资料统计分析，首先应该检验资料是否满足参数检验的前提条件，满足则选用单因素三水平设计定量资料方差分析，否则应选用此设计定量资料 Kruskal- Wallis H 检验。当三种基因型携带者的静息心率差别具有统计学意义时再进行两两基因型间的比较。

【案例 4-8】 原文目的探讨脊髓性肌萎缩(SMA)的临床表型与运动神经元生存基因(SMN)复制数变化之间是否存在相关性。对 57 例 SMA 患者(I 型 21 例、II 型 25 例、III 型 11 例)的 SMN₂ 基因复制数进行检测。数据分析用 ABI7000SDS 分析软件分析定量数据。结果显示 II 型 SMA 患者 SMN₂ 平均复制数明显高于 I 型($t = 4.24$, $P < 0.01$)，III 型 SMA 患者 SMN₂ 平均复制数明显高于 II 型($t = 2.44$, $P < 0.01$)，提示病情轻重与 SMN₂ 复制数有关，轻型患者可含有较多的 SMN₂ 基因复制。具体数据见表 4-4。

表 4-4 不同类型 SMA 患者 SMN₂ 复制数的定量分析

临 床 表 型	例 数	SMN ₂ 平均复制数($\bar{x} \pm s$)	t 值	P 值
I 型	21	1.926 \pm 0.460		

续表

临床表型	例数	SMN ₂ 平均复制数($\bar{x} \pm s$)	<i>t</i> 值	<i>P</i> 值
Ⅱ型	25	2.508 ± 0.460 ^a	4.24	0.001
Ⅲ型	11	2.876 ± 0.270 ^b	2.44	0.001

注: a, $P < 0.01$ (Ⅱ型与Ⅰ型 SMA 相比); b, $P < 0.01$ (Ⅲ型与Ⅱ型 SMA 相比)。

【案例辨析】从原文可以看出它主要是三种不同临床表型的 SMA 患者 SMN₂ 基因平均复制数的比较, 原文中并没有阐明采用何种统计分析方法, 从表 4-4 中看出它采用的是 *t* 检验, 但 *t* 检验只适用于不多于两组的数据间比较。

【案例释疑】该资料三种不同临床表型的 SMA 患者 SMN₂ 基因平均复制数的比较属于单因素三水平设计定量资料统计分析, 首先应该检验资料是否满足参数检验的前提条件, 满足则选用单因素三水平设计定量资料方差分析, 否则应选用此设计定量资料 Kruskal-Wallis H 检验。当三种不同临床表型的 SMA 患者 SMN₂ 基因平均复制数差别具有统计学意义时再进行两两临床表型间的比较。

4.2.2 误用 *t* 检验和单因素多水平设计定量资料方差分析处理多因素析因设计定量资料

【案例 4-9】原文目的探讨卵磷脂胆固醇酰基转移酶(LCAT)的基因 LCAT 608C/T 多态性在中国汉族人群中的分布及其与动脉粥样硬化性脑梗塞(ACI)的关联。研究对象为 150 例 ACI 患者和 122 例年龄、性别相匹配的正常对照。检测其 LCAT 608C/T 多态性。统计方法组间基因型和等位基因频率差异采用 χ^2 检验, 组间血脂比较采用 *t* 检验, 不同基因型亚组间血脂比较采用 ANOVA 方差分析, $P < 0.05$ 为差异有统计学意义。结果显示 ACI 组 CT 基因型频率、T 等位基因频率均显著高于对照组。ACI 和对照组中 608CC 亚组 HDL-C 水平均高于同组内 608CT 亚组。ACI 组和对照组各组内不同基因型亚组间血脂比较见表 4-5。

表 4-5 血脂 608CC 和 608CT 在 ACI 组和对照组两亚组间

血 脂	ACI(脑梗塞组)		对 照 组	
	608CC(<i>n</i> = 129)	608CT(<i>n</i> = 21)	608CC(<i>n</i> = 117)	608CT(<i>n</i> = 5)
TG(三酰甘油, mmol/L)	1.84 ± 0.85	1.88 ± 0.85	1.31 ± 0.46	1.35 ± 0.22
TC(总胆固醇, mmol/L)	4.75 ± 1.03	4.59 ± 1.08	4.31 ± 0.47	4.23 ± 0.21
HDL-C(高密度脂蛋白胆固醇, mmol/L)	1.15 ± 0.32	0.98 ± 0.30 *	1.57 ± 0.25	1.15 ± 0.16 [#]
LDL-C(低密度脂蛋白胆固醇, mmol/L)	2.68 ± 0.90	2.87 ± 1.08	2.22 ± 0.62	2.09 ± 0.75
apoA I (载脂蛋白 A I, g/L)	1.16 ± 0.37	1.06 ± 0.32	1.37 ± 0.63	1.34 ± 0.47
ApoB100(载脂蛋白 B100, g/L)	0.99 ± 0.31	1.05 ± 0.25	0.72 ± 0.14	0.75 ± 0.24

注: * 脑梗塞组中 608CC 与 608CT 亚组血脂比较, $F = 5.23$, $P = 0.024$; # 对照组 608CC 和 608CT 亚组血脂比较, $F = 6.489$, $P = 0.044$ 。

【案例辨析】从原文可以看出 ACI 组和对照组各组内不同基因型亚组间血脂比较采用的是单因素方差分析的方法, 而本资料显然涉及两个因素: 一个为分组因素, 分为两个水平(ACI 组和对照组); 另一个因素是不同的基因型亚组, 也分为两个水平(608CC 和 608CT)。此设计类型是两因素析因设计或 2×2 析因设计, 原文采用单因素方差分析处理该定量资料是错误的。

【案例释疑】该资料设计类型为两因素析因设计, 若资料满足参数检验的前提条件, 则采用两因素析因设计定量资料方差分析。若不满足参数检验的前提条件, 应进行适当的变量变换, 当变换后的资料满足参数检验的前提条件时, 对变换后的资料采用两因素析因设计定量资料方差分析来进行处理。

值得注意的是, 本例所涉及的两个试验因素对某些结果的影响可能是不平等的, 若果真如

此,应以专业知识为依据,如组别之间的差别大于两种基因型之间的差别,应判为两因素嵌套或系统分组设计。

【案例 4-10】 原文目的探讨白细胞介素 6(IL6) 基因 -572C/G 多态性在中国人群中的分布频率与心肌梗死(MI)易感性的关系,对 MI 患者冠脉病变程度的影响以及初步对该位点基因变异进行功能性分析。研究对象为 232 例 MI 患者和 260 例正常对照。检测其 IL6 基因 -572C/G 多态性。统计方法组间一般临床资料的比较采用 t 检验或 χ^2 检验,组间人群基因型与等位基因频率的比较均采用 χ^2 检验。结果显示两组人群基因型、等位基因频率差异存在统计学意义。不同基因型间血脂、脂蛋白及载脂蛋白水平的比较得 MI 组基因型间 TG 水平差异有统计学意义($P<0.05$)。具体数据见表 4-6。

表 4-6 血脂、脂蛋白及载脂蛋白水平在不同基因型人群间的比较

变 量	$\bar{X} \pm s$			
	¥: CC	CG + GG	\$: CC	CG + GG
TC(总胆固醇, mmol/L)	4.40 ± 0.59	4.43 ± 0.51	4.94 ± 1.15	5.02 ± 1.24
TG(三酰甘油, mmol/L)	1.14 ± 0.46	1.20 ± 0.53	1.82 ± 1.91	2.56 ± 1.88 *
HDL-C(高密度脂蛋白胆固醇, mmol/L)	1.24 ± 0.70	1.13 ± 0.68	1.31 ± 0.55	1.23 ± 0.54
LDL-C(低密度脂蛋白胆固醇, mmol/L)	2.20 ± 0.92	2.42 ± 0.85	3.47 ± 1.41	3.66 ± 1.78
apoA I(载脂蛋白 A I, g/L)	1.23 ± 0.39	1.15 ± 0.30	1.21 ± 0.34	1.15 ± 0.32
apoB(载脂蛋白 B, g/L)	1.18 ± 0.22	1.21 ± 0.26	1.19 ± 0.25	1.23 ± 0.22

注: * 与 MI 组 CC 基因型比较 $t=3.92$, $P<0.05$; ¥代表“对照组基因型”;\$:代码“MI 组基因型”。

【案例辨析】 从原文可以看出不同基因型间血脂、脂蛋白及载脂蛋白水平的比较采用的是 t 检验,而本资料显然涉及两个因素:一个为分组因素,分为两个水平(MI 组和对照组);另一个因素是不同的基因型,也分为两个水平(CC 和 CG + GG)。此设计类型是两因素析因设计或 2×2 析因设计,原文采用 t 检验处理该定量资料是错误的。

【案例释疑】 该资料设计类型为两因素析因设计,首先考虑采用两因素析因设计定量资料方差分析,但需要满足参数检验的前提条件。否则应进行变量变换,如变换后的资料满足参数检验的前提条件时,则可对变换后的资料采用两因素析因设计定量资料方差分析来进行处理。

值得注意的是,本例所涉及的两个试验因素对表中前 4 项指标的影响可能是不平等的,若果真如此,应以专业知识为依据,如组别之间的差别大于两种基因型之间的差别,应判为两因素嵌套或系统分组设计。

4.2.3 误用 t 检验和单因素方差分析处理多因素重复测量设计定量资料

【案例 4-11】 原文目的检测首发精神分裂症患者外周血淋巴细胞中神经调节素 1(NRG1) 基因 mRNA 表达水平及其在治疗后的变化,分析该基因的表达和精神分裂症之间的关联。研究对象选取了精神分裂症患者 80 例,同胞对照组 37 例,正常对照 83 例。应用 RT-PCT 方法测定患者服抗精神药物前、服药后第 1、2、3、4 周的外周血 NRG1 mRNA 的表达水平,同时测定两个对照组 NRG1 mRNA 的表达水平。应用独立样本 t 检验比较病例组经利培酮和奎硫平治疗后组间不同时间点 NRG1 mRNA 相对表达量,结果显示两组患者服药后不同时间点 NRG1 mRNA 相对表达量差异均无统计学意义($P>0.05$),数据见表 4-7。

【案例辨析】 原文比较利培酮治疗组和奎硫平治疗组间不同时间点 NRG1 mRNA 相对表达量用了独立样本 t 检验,这很明显是错误的,因为同一个研究对象不同时间点的表达量是不独立的,不能把各个时间点割裂来分析。从研究设计中可以看到测定的是同一组研究对象服药前、服药后

第1、2、3、4周的NRG1 mRNA的表达水平,这个属于典型的重复测量设计。从表2-7中很容易看出其对应的试验设计类型为“含有一个重复测量的两因素设计定量资料”。一个因素为分组因素,有两个水平(利培酮治疗组和奎硫平治疗组);另外一个因素为时间因素,这是一个与重复测量有关的试验因素,有5个水平(0、1、2、3、4周)。很显然用 t 检验处理两因素重复测量设计的定量资料是错误的。

表4-7 利培酮组和奎硫平组服药后不同时间NRG1 mRNA相对表达量比较

时 间	利培酮组($n=47$)	奎硫平组($n=33$)	t 值	P 值
0 周	2.20 ± 1.02	2.11 ± 0.82	0.432	0.667
1 周	2.47 ± 0.68	2.44 ± 0.63	0.201	0.847
2 周	3.17 ± 1.27	3.07 ± 1.15	0.374	0.709
3 周	3.49 ± 1.03	3.52 ± 1.17	-0.147	0.883
4 周	3.75 ± 1.54	3.68 ± 1.67	0.197	0.844

【案例释疑】该资料设计类型为“具有一个重复测量的两因素设计”,若资料满足参数检验的前提条件,则采用此设计定量资料方差分析。若资料不满足参数检验的前提条件,应做适当的变量变换,当变换后的资料满足参数检验的前提条件后,对变换后的资料采用此设计定量资料方差分析处理。

【案例4-12】原文目的通过对不同类型高苯丙氨酸血症(HPA)临床特点的分析,探讨我国南、北方四氢生物蝶呤(BH_4)反应性苯丙氨酸羟化酶(PAH)缺乏症患者对 BH_4 的反应性。选取108例HPA患儿,进行口服 BH_4 负荷试验,在服药前以及服药后2h、4h、8h、24h分别留血,测定血苯丙氨酸(Phe)浓度,并计算出不同时间点血Phe浓度与服药前相比下降的百分数。108例HPA患儿中确诊 BH_4 反应性PAH缺乏症36人,根据其父母祖籍以长江为界分为南方、北方两组。应用SPSS 11.5软件比较南、北两组 BH_4 反应性PAH缺乏症患者在 BH_4 负荷试验各时间点血Phe浓度较服药前下降的百分率。结果显示南、北方两组在服药2h、4h、8h、24h时血Phe浓度下降的比率差异无统计学意义($P>0.05$)。具体数据见表4-8。

表4-8 南、北方组 BH_4 反应性PAH缺乏症 BH_4 负荷试验中各时间点血Phe浓度下降百分数的比较

时间(h)	血Phe浓度下降的百分数(%)	
	南方患者($n=23$)	北方患者($n=13$)
2	23.72 ± 15.81	15.82 ± 14.96
4	36.20 ± 25.37	26.18 ± 22.80
8	51.19 ± 21.77	45.81 ± 25.83
24	69.56 ± 19.76	57.88 ± 18.88

注:各组比较均 $P>0.05$ 。

【案例辨析】原文比较南、北两组 BH_4 反应性PAH缺乏症患者在 BH_4 负荷试验各时间点血Phe浓度较服药前下降的百分率的差异,虽然原文中并未说明采用何种统计分析方法,但从表4-8中可以看出,也是进行的各个时间点两组间的比较,或者是独立样本 t 检验,或者单因素方差分析。这两种方法均不适合处理该类型下的定量资料。从原文看出HPA患儿在服药前以及服药后2h、4h、8h、24h分别留血测定Phe浓度,说明该资料也属于重复测量设计。一个因素为分组因素,有两个水平(南方组和北方组);另外一个因素为时间因素,这是一个与重复测量有关的试验因素,有四个水平(2h、4h、8h、24h)。所以说采用 t 检验或单因素方差分析处理这种重复测量设计定量资料是错误的。

比较利培酮治疗组和奎硫平治疗组间不同时间点NRG1 mRNA相对表达量用了独立样本 t 检

验,这很明显是错误的,因为同一个研究对象不同时间点的表达量是不独立的,不能把各个时间点割裂开来再分析。从研究设计中可以看到测定的是同一组研究对象服药前、服药后第1、2、3、4周的NRG1 mRNA的表达水平,这个属于典型的重复测量设计。从表2-6中很容易看出其对应的试验设计类型为“含有一个重复测量的两因素设计定量资料”。一个因素为分组因素,有两个水平(利培酮治疗组和奎硫平治疗组);另外一个因素为时间因素,这是一个与重复测量有关的试验因素,有5个水平(0、1、2、3、4周)。很显然用 t 检验处理此设计定量资料是错误的。

【案例释疑】 该资料设计类型为“具有一个重复测量的两因素设计定量资料”,首选此设计定量资料方差分析,但它要求资料满足参数检验的前提条件。若资料不满足参数检验的前提条件,则做适当的变量变换后采用此设计定量资料方差分析处理。

4.2.4 本节相关知识简要回顾与小结

在选择统计分析方法前,首先要清楚研究目的是什么,要判断资料的类型是什么,资料所取自的设计类型是什么,同时还要考察是否满足相应方法的应用条件,综合判断以后才能对统计分析方法作出合理的选择。例如,用参数检验处理定量资料时,通常对资料的正态性和方差齐性有要求。设计类型有成组设计、配对设计、随机区组设计、析因设计、正交设计、重复测量设计等。不同的设计类型对应的统计分析方法不一样,所以在应用时一定要注意。

4.3 定性资料统计分析中存在的问题

4.3.1 多重比较的校正问题

【案例4-13】 原文目的为了了解中国汉族人群DG-SIGN和DC-SIGNR基因颈区重复序列的遗传多态性分布,获得相应位点的汉族人群的遗传学数据。选取520例中国汉族正常人群,检测DG-SIGN和DC-SIGNR基因颈区重复序列的遗传多态性。计算等位基因频率和基因型频率,不同种族间等位基因及基因型频率的比较用 χ^2 检验。中国汉族人群DC-SIGNR基因型频率和等位基因频率与美国白种人群比较结果显示4、5、6、7、9等位基因分布频率在中国汉族人群(520名)和美国白种人(698名)中的差异有统计学意义($P<0.05$);6/5、7/4、7/5、7/6、7/7、9/5、9/7、9/9基因型在两组人群中差异也有统计学意义($P<0.05$)。具体数据见表4-9和表4-10。

表4-9 中国汉族人和美国白人DC-SIGNR等位基因频率分布

DC-SIGNR 等位基因	等位基因频率	
	中国汉族人	美国白人
4 *	0.0096	0.0201
5 *	0.1500	0.2493
6 *	0.0462	0.1483
7 *	0.6500	0.5594
8	0.0048	0.0014
9 *	0.1394	0.0215
总计	1.0000	1.0000

注: * 为 χ^2 检验差异有统计学意义的等位基因频率分布。

【案例辨析】 原文比较了中国汉族人群和美国白种人在DC-SIGNR基因型频率和等位基因频率上的差异,文中显示直接采用了 χ^2 检验的方法。不管是两种人群间等位基因频率的比较还是两

种人群间基因型频率的比较,任何两个等位基因或基因型都进行了一次统计学检验,多次检验会导致假阳性率迅速增大,从而使得结论不可信,因此需要对分析结果进行统计学校正。

表 4-10 中国汉族人和美国白人 DC-SIGNR 基因型分布

DC-SIGNR 基因型	数 量	基 因 频 率	数 量	基 因 频 率
	中国汉族人		美国白人	
4/4	2	0.0038	2	0.0029
5/4	—	—	6	0.0086
5/5	23	0.0442	49	0.0702
6/4	2	0.0038	3	0.0043
6/5 *	8	0.0154	51	0.0731
6/6	6	0.0115	18	0.0258
7/4 *	1	0.0019	15	0.0215
7/5 *	90	0.1731	189	0.2708
7/6 *	22	0.0423	116	0.1662
7/7 *	233	0.4481	217	0.3109
8/5	1	0.0019	—	—
8/8	2	0.0038	—	—
8/6	—	—	1	0.0014
8/7	—	—	1	0.0014
9/4	2	0.0058	—	—
9/5 *	12	0.0212	4	0.0057
9/6	4	0.0077	—	—
9/7 *	97	0.1865	26	0.0372
9/9 *	15	0.0288	—	—
总计	520	1.0000	698	1.0000

注: * 为 χ^2 检验差异有统计学意义的基因型。

【案例释疑】常用的多重比较校正方法是 Bonferroni 法校正 P 值,即 $P_c = \text{原 } P \text{ 值} \times \text{比较的次数}$, $P_c \leq 0.05$ 为差异有统计学意义。两种人群间等位基因频率比较的 $P_c = \text{原 } P \text{ 值} \times 6$ (等位基因数); 两种人群间基因型频率比较的 $P_c = \text{原 } P \text{ 值} \times 19$ (基因型数)。重新计算得到的 P_c 与 0.05 比较再看差异是否有统计学意义。从表 4-11 可见,校正 P 值后 DC-SIGNR 等位基因 4 的频率在中国汉族人和美国白人间差异无统计学意义。从表 4-12 可见校正 P 值后 DC-SIGNR 基因型 7/4 和 9/5 的频率在中国汉族人和美国白人间差异无统计学意义。

表 4-11 中国汉族人和美国白人 DC-SIGNR 等位基因频率分布比较

DC- SIGNR 等位基因	例 数	频 率	例 数	频 率	χ^2	P	P_c
	中国汉族人($n=520$)		美国白人($n=698$)				
4	10	0.0096	28	0.0201	4.232	0.040	0.240
5	156	0.1500	348	0.2493	35.803	2.18×4^{-9}	1.31×4^{-8}
6	48	0.0462	207	0.1483	66.326	3.82×4^{-16}	2.29×4^{-15}
7	676	0.6500	781	0.5594	20.327	6.53×4^{-6}	3.92×4^{-5}
8	5	0.0048	2	0.0014	1.338	0.247	1.000
9	145	0.1394	30	0.0215	124.317	7.18×4^{-29}	4.31×4^{-28}
总计	1040	1.0000	1396	1.0000			

表 4-12 中国汉族人和美国白人 DC-SIGNR 基因型分布比较

DC-SIGNR 基因型	例 数	频 率	例 数	频 率	χ^2	P	P_c
	中国汉族人 ($n=520$)		美国白人 ($n=698$)				
4/4	2	0.0038	2	0.0029	0	1.000	1.000
5/4	—	—	6	0.0086			
5/5	23	0.0442	49	0.0702	3.613	0.057	1.000
6/4	2	0.0038	3	0.0043	0	1.000	1.000
6/5	8	0.0154	51	0.0731	21.510	3.52×4^{-6}	6.69×4^{-5}
6/6	6	0.0115	18	0.0258	3.132	0.077	1.000
7/4	1	0.0019	15	0.0215	8.801	0.003	0.057
7/5	90	0.1731	189	0.2708	16.106	5.99×4^{-5}	0.001138
7/6	22	0.0423	116	0.1662	45.521	1.51×4^{-11}	2.87×4^{-10}
7/7	233	0.4481	217	0.3109	24.075	9.26×4^{-7}	1.76×4^{-5}
8/5	1	0.0019	—	—			
8/8	2	0.0038	—	—			
8/6	—	—	1	0.0014			
8/7	—	—	1	0.0014			
9/4	2	0.0058	—	—			
9/5	12	0.0212	4	0.0057	6.917	0.009	0.171
9/6	4	0.0077	—	—			
9/7	97	0.1865	26	0.0372	73.155	1.20×4^{-17}	2.28×4^{-16}
9/9	15	0.0288	—	—			
总计	520	1.0000	698	1.0000			

【案例 4-14】 原文目的分析兰州地区汉族人群 HLA-A、B 和 DRB1 位点等位基因多态性特点。方法是选取了兰州地区 200 名健康无血缘关系的汉族个体 HLA-A、B 和 DRB1 基因进行分型，并与西北、北方和南方汉族、西北回族、维吾尔族和藏族人群进行比较。统计学分析直接计算基因频率，组间频率比较用 χ^2 检验。HLA- DRB1 等位基因频率与南、北汉族、西北汉族和西北少数民族比较，与西北回族差异最小，其次为北方汉族、西北汉族和南方汉族，与西北维吾尔族和藏族差异有统计学意义。具体数据见表 4-13。

表 4-13 兰州地区汉族与不同人群 HLA-DRB1 等位基因频率比较

等 位 基 因	兰州汉族 ($n=200$)	北方汉族 ($n=2258$)	南方汉族 ($n=4707$)	西北汉族 ($n=183$)	西北回族 ($n=169$)	西北维族 ($n=200$)	西北藏族 ($n=188$)
DRB1 * 01	2.75	2.69	0.99 Δ	2.76	4.10	11.25 Δ	0.80 Δ
DRB1 * 03	6.00	1.20 Δ	2.47 Δ	4.64	4.37	2.00 Δ	0.80 Δ
DRB1 * 04	10.75	11.46	11.69	13.93	10.38	20.50 Δ	24.47 Δ
DRB1 * 07	9.75	10.84	4.79 Δ	13.10	12.57	23.75 Δ	5.85 $\#$
DRB1 * 08	7.50	6.83	8.14	3.83 $\#$	4.64	0 Δ	6.38
DRB1 * 09	14.00	13.59	15.89	12.02	12.02	1.25 Δ	9.30 $\#$
DRB1 * 10	3.00	1.90	1.77	2.73	2.19	0.25 Δ	1.10
DRB1 * 11	10.00	6.26	6.30	5.19	7.10	6.75	13.03
DRB1 * 12	11.50	10.94	12.94	12.57	8.74	6.75 $\#$	13.30
DRB1 * 13	3.25	6.10 $\#$	4.73	8.74 Δ	5.46	10.00 Δ	7.71 Δ
DRB1 * 14	7.25	6.00	7.05	3.55 $\#$	7.10	7.50	8.51
DRB1 * 15	11.75	16.19	14.08	13.66	10.93	8.00	5.32 Δ
DRB1 * 16	2.50	2.31	4.71 $\#$	3.28	2.46	2.00	15.96 Δ

注：#， $P < 0.05$ ； Δ ， $P < 0.01$ 。

【案例辨析】 原文比较了兰州汉族与西北、北方和南方汉族、西北回族、维吾尔族和藏族人群在 HLA-DRB1 等位基因频率上的差异, 原文中写明直接采用了 χ^2 检验的方法。由于 HLA-DRB1 等位基因数较多, 共 13 个等位基因, 所以任何两个人群间比较都要进行 13 次 χ^2 检验, 多次检验会导致假阳性率迅速增大, 从而使得结论不可信, 因此需要对分析结果进行统计学校正。

【案例释疑】 多重比较用 Bonferroni 法校正 P 值, 即 $P_c = \text{原 } P \text{ 值} \times \text{比较的次数}$, $P_c \leq 0.05$ 为差异有统计学意义。任何两个人群间 HLA-DRB1 等位基因频率比较的 P 值都需要校正, $P_c = \text{原 } P \text{ 值} \times 13$ (等位基因数)。重新计算得到的 P_c 与 0.05 比较再看差异是否有统计学意义。这里以兰州地区汉族和西北汉族的比较为例, 从表 4-14 可见, 原文中等位基因 DRB1 * 08、DRB1 * 11、DRB1 * 14 在兰州地区汉族和西北汉族之间差异有统计学意义, P 值经校正后均无统计学意义, 最终可知兰州地区汉族和西北汉族仅在 HLA-DRB1 * 13 等位基因差异有统计学意义。

表 4-14 兰州地区汉族与西北汉族 HLA-DRB1 等位基因频率比较

HLA-DRB1 等位基因	例 数	频率(%)	例 数	频率(%)	χ^2	P	P_c
	兰州汉族($n=200$)		西北汉族($n=183$)				
DRB1 * 01	11	2.75	10	2.76	2.26×4^{-4}	0.988	1.000
DRB1 * 03	24	6.00	17	4.64	0.693	0.405	1.000
DRB1 * 04	43	10.75	51	13.93	1.800	0.180	1.000
DRB1 * 07	39	9.75	48	13.10	2.149	0.143	1.000
DRB1 * 08	30	7.50	14	3.83	4.767	0.029	0.377
DRB1 * 09	56	14.00	44	12.02	0.659	0.417	1.000
DRB1 * 10	12	3.00	10	2.73	0.049	0.825	1.000
DRB1 * 11	40	10.00	19	5.19	6.217	0.013	0.169
DRB1 * 12	46	11.50	46	12.57	0.206	0.650	1.000
DRB1 * 13	13	3.25	32	8.74	10.430	0.001	0.013
DRB1 * 14	29	7.25	13	3.55	5.044	0.025	0.325
DRB1 * 15	47	11.75	50	13.66	0.631	0.427	1.000
DRB1 * 16	10	2.50	12	3.28	0.415	0.519	1.000
总计	400	1.0000	366	1.0000			

【案例 4-15】 原文目的调查辽宁汉族人群 HLA-B 等位基因的遗传多态性。选取辽宁 8962 名健康无关汉族人进行 HLA-B 等位基因分型, 计算 HLA-B 等位基因频率, 不同人群间等位基因频率比较采用 χ^2 检验。结果显示辽宁汉族人群共检测出 HLA-B 等位基因 34 种。与不同人群 HLA-B 等位基因频率比较, 辽宁汉族与南方汉族相比除 HLA-B * 15、* 45、* 47、* 53、* 54、* 59 和 * 81, 其余差异有统计学意义 ($P < 0.05$); 与北方汉族相比除 HLA-B * 7、* 27、* 35、* 41、* 44、* 46、* 48、* 57 和 * 58, 其余差异无统计学意义 ($P > 0.05$)。

【案例辨析】 原文将辽宁汉族人群 HLA-B 等位基因频率分别与南方汉族、北方汉族比较, 直接采用了 χ^2 检验, 因等位基因有 34 种, 所以增大了犯 I 型错误的概率, 因此需要对分析结果进行统计学校正。

【案例释疑】 应用 Bonferroni 法校正 P 值, 即 $P_c = \text{原 } P \text{ 值} \times 34$ (等位基因数), $P_c \leq 0.05$ 为差异有统计学意义。

【案例 4-16】 原文研究目的探讨皖籍汉族人群 MICA 基因第 2、3、4 外显子的多态性及其与 HLA-B 抗原的连锁不平衡在强直性脊柱炎 (AS) 发病中的作用。研究对象选取 56 例 AS 患者和 112 名正常对照人群进行 MICA 基因第 2、3、4 外显子的多态性和 HLA-B 抗原检测。病例组和正常对照组间 MICA 等位基因频率差异的统计学意义采用 χ^2 检验。 χ^2 检验结果显示, AS 患者和正常对照组

MICA * 007 等位基因的分布差异有统计学意义($P < 0.05$)，两组间其他 MICA 等位基因分布的差异无统计学意义($P > 0.05$)，具体见表 4-15。

表 4-15 MICA 等位基因在 AS 患者和正常对照组的分布

MICA 等位基因	AS 患者组 (n, %)	正常对照组 (n, %)	χ^2 值	P 值	95% 可信区间
MICA * 008	36(32.14)	68(30.36)	0.20	0.65	1.06(0.83 – 1.35)
MICA * 010	18(16.07)	38(16.96)	0.05	0.82	0.95(0.60 – 1.50)
MICA * 002	11(9.82)	31(13.84)	1.29	0.26	0.71(0.39 – 1.30)
MICA * 009	10(8.93)	27(12.05)	0.85	0.36	0.74(0.39 – 1.42)
MICA * 004	8(7.14)	20(8.93)	0.34	0.56	0.80(0.38 – 1.70)
MICA * 012	6(5.36)	19(8.48)	1.15	0.28	0.63(0.27 – 1.49)
MICA * 007	20(17.86)	16(7.14)	10.18	0.001	2.50(1.41 – 4.44)
MICA * 005	3(2.68)	5(2.23)	Fisher	0.07	1.20(0.30 – 4.84)

【案例辨析】 原文 AS 患者和正常对照组 MICA * 007 等位基因的分布差异比较采用 χ^2 检验，等位基因数为 8，所以要进行 8 次 χ^2 检验，增大犯 I 型错误的概率，因此要对 P 值进行 Bonferroni 校正。

【案例释疑】 应用 Bonferroni 法校正 P 值，即 $P_c = \text{原 } P \text{ 值} \times 8$ (等位基因数)， $P_c \leq 0.05$ 为差异有统计学意义。重新计算结果见表 4-16。校正后 MICA * 007 等位基因 P 值为 0.024，差异仍有统计学意义。

表 4-16 MICA 等位基因在 AS 患者和正常对照组的分布

MICA 等位基因	AS 患者组 (n, %)	正常对照组 (n, %)	χ^2	P	P_c
MICA * 008	36(32.14)	68(30.36)	0.111	0.739	1.000
MICA * 010	18(16.07)	38(16.96)	0.043	0.836	1.000
MICA * 002	11(9.82)	31(13.84)	1.102	0.294	1.000
MICA * 009	10(8.93)	27(12.05)	0.744	0.388	1.000
MICA * 004	8(7.14)	20(8.93)	0.312	0.577	1.000
MICA * 012	6(5.36)	19(8.48)	1.059	0.303	1.000
MICA * 007	20(17.86)	16(7.14)	8.960	0.003	0.024
MICA * 005	3(2.68)	5(2.23)	0.000	1.000	1.000

【案例 4-17】 原文研究目的探讨特发性扩张型心肌病 (IDC) 患者细胞毒性 T 淋巴细胞相关抗原 -4 (CTLA-4) 表达状况及由 CTLA-4 基因启动子区单核苷酸多态性导致的不同遗传易感性机制。研究对象选取 151 例 IDC 患者和 120 名正常健康人，检测其 CTLA-4 基因启动子区 -1772、-1661 及 -318 位点的多态性。直接计数法计算基因频率和基因型频率，采用 χ^2 检验进行 H - W 平衡吻合度检验。结果显示对照组 CTLA-4 基因 -1661 位点 GG 基因型频率显著高于 IDC 组 ($P = 0.018$ ，OR = 0.065)，且 G 等位基因频率增高 ($P = 0.012$ ，OR = 0.356)。具体数据见表 4-17。

表 4-17 CTLA-4 基因 -1661 位点基因型及等位基因频率分布

分 组	基因型频率 (%)			等位基因频率 (%)	
	AA	AG	GG	A	G
正常对照组 (n = 120)	52(43.10)	26(21.54)	42(35.36) *	130(54.17)	110(45.83) #
IDC 组 (n = 151)	88(58.28)	42(27.81)	20(13.24)	219(72.52)	83(27.48)

注：*，与 IDC 组比较， $P = 0.018$ ，OR = 0.065，95% CI (0.006 – 0.191)；#，与 IDC 组比较， $P = 0.012$ ，OR = 0.356，95% CI (0.091 – 0.645)。

【案例辨析】 对照组和 IDC 组 CTLA-4 基因 -1661 位点 GG 基因型频率的比较应该属于两组

构成比的比较, 首先应该整体检验两组构成是否有差异, 而原文并未给出两组整体检验的结果, 而是直接给出对照组和 IDC 组 GG 基因型频率的较的结果, 且计算结果是错误的。一般进行整体检验构成差异有统计学意义的话可进一步做多重比较, 但检验水准需要 Bonferroni 校正, 因为构成比合计是固定的, 所以校正数为两两比较次数 - 1。当检验某一基因型时, 该基因型频数不变, 其他两个基因型合并。

【案例释疑】 整体检验两组构成比差异有统计学意义($\chi^2 = 17.71$, $P = 0.000$)。多重比较结果见表 4-18。

表 4-18 CTLA-4 基因 -1661 位点基因型频率分布在对照组和 IDC 组间的比较

基 因 型	频率 (%)		χ^2	原 P 值	校正 P 值
	IDC 组 (n=151)	正常对照组 (n=120)			
AA	88 (58.28)	52 (43.10)	6.278	0.012	0.024
AG	42 (27.81)	26 (21.54)	1.419	0.234	0.468
GG	20 (13.24)	42 (35.36)	17.692	2.6×4^{-5}	5.2×4^{-5}

整体检验: $\chi^2 = 17.71$, $P = 0.000$ 。

从表 4-18 可知 IDC 组 GG 基因型频率低于对照组, 差异有统计学意义($P < 0.05$); IDC 组 AA 基因型频率高于对照组, 差异有统计学意义($P < 0.05$)。

4.3.2 分层分析缺少 CMH χ^2 检验

【案例 4-18】 原文研究目的探讨 X 线修复交叉互补组 1 基因(XRCC1)、8-羟基鸟嘌呤修复酶基因(hOGG1)多态性与喉癌遗传易感性的关系。采用病例对照设计, 检测了 72 例经病理确诊的喉癌患者和随机抽样的 72 例无肿瘤、无遗传病对照者 XRCC1-Arg399Gln、hOGG1-Ser326Cys 多态性。方法采用 χ^2 检验, 计算发病风险比值比及其 95% 可信区间。对 XRCC1-Arg399Gln、hOGG1-Ser326Cys 各基因型分布频率进行分析, 表 4-19 结果显示病例组与对照组之间基因型构成比差异存在统计学意义, XRCC1($\chi^2 = 12.37$, $P = 0.002$)、hOGG1($\chi^2 = 9.62$, $P = 0.008$)。

表 4-19 XRCC1、hOGG1 基因多态性与喉癌易感性的关系

多 态 位 点	基 因 型	对照组 (n, %)	喉癌组 (n, %)	OR (95% CI)
XRCC1 399	Arg/Arg	43 (59.7)	22 (30.6)	1.00
	Arg/Gln	22 (30.6)	38 (52.8)	3.38 (1.62 - 7.04) ^a
	Gln/Gln	7 (9.7)	12 (16.7)	3.35 (1.16 - 9.71) ^b
	Arg/Gln + Gln/Gln	29 (40.3)	50 (69.4)	3.37 (1.69 - 6.70) ^a
hOGG1 326	Ser/Ser	50 (69.4)	34 (47.2)	1.00
	Ser/Cys	22 (30.6)	34 (47.2)	2.27 (1.14 - 4.53) ^b
	Cys/Cys	0 (0)	4 (5.6)	
	Ser/Cys + Cys/Cys	22 (30.6)	38 (52.8)	2.54 (1.28 - 5.03) ^b

注: 在喉癌组与对照组之间, 杂合型、突变型与野生型的频率差异经 χ^2 检验, a: $P < 0.01$, b: $P < 0.05$ 。说明由此而计算出来的发病风险比值比有统计学意义。

危险度分析显示, 携带 XRCC1 - Arg399Arg/Gln 或 Gln/Gln 基因型的个体较携带 Arg/Arg 基因型个体喉癌的发病风险平均升高了 3.37 倍; 携带 hOGG1 - Ser326Ser/Cys 或 Cys/Cys 基因型个体喉癌的发病风险较携带 Ser/Ser 基因型的个体平均升高了 2.54 倍。

吸烟与 XRCC1、hOGG1 基因的交互作用分析显示: 在吸烟组或不吸烟组, XRCC1 - 399 杂合型及突变型 (Arg/Gln + Gln/Gln) 和 hOGG1 - 326 杂合型及突变型 (Ser/Cys + Cys/Cys) 在病例组比例

均高于对照组, OR 值分别为 4.07 或 3.07 和 2.32 或 3.41; 分层分析的同质性检验显示不同层间(吸烟组与不吸烟组)差异无统计学意义($XRCC1: \chi^2_{H1} = 0.15, P = 0.69$; $hOGG1: \chi^2_{H2} = 0.28, P = 0.60$), 见表 4-20。

表 4-20 吸烟与 XRCC1、hOGG1 基因交互作用分析

基 因 型	吸烟组 (n)		OR(95%CI)	不吸烟组 (n)		OR(95%CI)
	对 照 组	病 例 组		对 照 组	病 例 组	
XRCC1 - 399 ^a						
Arg/Arg	24	9	1.00	19	13	1.00
Arg/Gln + Gln/Gln	19	29	4.07	10	21	3.07
		1.56 - 10.63			1.09 - 8.61	
hOGG1 - 326 ^b						
Ser/Ser	27	16	1.00	23	18	1.00
Ser/Cys + Cys/Cys	16	22	2.32	6	16	3.41
		0.95 - 5.67			1.11 - 10.47	

注: 吸烟与不吸烟组的 OR 组同质性 χ^2 检验, a: $\chi^2_{H1} = 0.15, P > 0.05$; b: $\chi^2_{H2} = 0.28, P > 0.05$ 。

【案例辨析】 原文首先在未分层前进行了 XRCC1、hOGG1 基因与喉癌的关联性分析, 得到了 OR 值, 之后为了评估吸烟对 XRCC1、hOGG1 基因和喉癌之间关系的可能影响, 资料按吸烟分层后进行分析, 同质性检验结果显示各层 OR 值是一致的($P > 0.05$), 也就是说吸烟与 XRCC1、hOGG1 基因之间不存在交互作用。这时可以把分层资料综合起来估计合并的 OR 值, 如 Mantel-Haenszel OR 值, 用合并 OR_{MH} 及其可信区间估计 XRCC1、hOGG1 基因和喉癌的关系。如果同质性检验结果显示各层 OR 值差异有统计学意义($P < 0.05$), 说明吸烟与 XRCC1、hOGG1 基因之间存在交互作用, 这时需要分别报道吸烟组的 OR 值及可信区间和不吸烟组的 OR 值及其可信区间来准确描述 XRCC1、hOGG1 基因和喉癌的关系。

【案例释疑】 经 CMH χ^2 检验得 XRCC1 基因: $\chi^2_{CMH} = 12.979, P = 0.0003, OR_{MH} = 3.575$ 。hOGG1 基因: $\chi^2_{CMH} = 7.917, P = 0.0049, OR_{MH} = 2.703$ 。携带 XRCC1 - Arg399Arg/Gln 或 Gln/Gln 基因型的个体发生喉癌的风险是携带 Arg/Arg 基因型个体的 3.575 倍; 携带 hOGG1 - Ser326Ser/Cys 或 Cys/Cys 基因型个体发生喉癌的风险是携带 Ser/Ser 基因型个体的 2.703 倍。

4.3.3 误用 χ^2 检验处理结果变量为有序变量的单向有序列联表资料

【案例 4-19】 原文研究目的探讨 L-选择素基因 P213S 多态性是否与冠状动脉粥样硬化性心脏病的关联。研究对象选取 212 例经冠状动脉造影确诊的冠心病患者和 230 名正常对照者进行研究。统计方法: 组间基因型与等位基因频率比较用 χ^2 检验。为了研究 L-选择素基因多态性与冠心病严重程度的关系, 根据冠状动脉造影结果将冠心病组分为单支血管病变组、双支血管病变组和三支血管病变组, L-选择素基因多态性与冠脉血管病变程度的关系见表 4-21。结果显示 L-选择素基因多态性与病变血管支数及疾病严重程度无相关性(。

表 4-21 冠状动脉血管病变程度与 L-选择素 P213S 多态性的关系

病 变 支 数	例 数	基因型频率 (%)		P 值
		PP	PS + SS	
1	107	60 (48.39)	47 (53.41)	0.366
2	64	36 (29.03)	28 (31.82)	
3	41	28 (22.58)	13 (14.77)	

【案例辨析】 原文目的研究 L- 选择素基因 P213S 多态性与冠状动脉血管病变程度的关系，从表中可见结果变量病变支数是等级变量，原文中采用 χ^2 检验是错误的。

【案例释疑】 可将不同基因型作为分组变量，病变支数为结果变量，目的是比较不同基因多态性间冠状动脉血管病变程度是否有差异，因结果变量是有序变量，所以资料属于单向有序 $2 \times C$ 列联表资料，应采用成组设计有序资料 Wilcoxon 秩和检验来处理，得 $P = 0.283$ ，按照 0.05 检验水准，不同基因型间冠状动脉血管病变程度差异无统计学意义。

4.3.4 本节相关知识简要回顾与小结

定性资料统计分析方法最常用的是 χ^2 检验，但是 χ^2 检验的错误应用也是最多的。对于不同类型的列联表资料统计分析方法不同，而且一定要结合研究目的选择方法。 χ^2 检验本身也有应用条件，有校正的 χ^2 检验，有 Fisher 确切概率法作补充。除此之外还有进行趋势分析的趋势 χ^2 检验和分层分析的 CMH χ^2 检验。同时要注意多重比较的校正问题。对于等级资料，常选用秩和检验来推断等级强度的差别，若用 χ^2 检验，则只能推断构成比的差别。如果小样本定量资料不满足正态性或方差齐性，这时也应用秩和检验来分析。

4.4 多变量回归分析方面存在的问题

4.4.1 多重线性回归方面存在的问题

【案例 4-20】 原文研究目的探讨纤维蛋白原 (Fg) β 基因启动子区 -148C/T 多态性、血浆 Fg 水平与急性动脉粥样硬化性脑梗死的关系。研究对象选择 151 例脑梗死患者和 101 名健康对照进行。统计方法：计量资料采用 t 检验 (由于血浆 Fg 浓度呈偏态分布，故 t 检验前将数据做对数转换；由于 TT 基因型例数太少，将 TT 型并入 CT 型与 CC 型进行比较)， $P < 0.05$ 为差异有统计学意义。Fg β -148C/T 基因多态性与血浆 Fg 水平：病例组和对照组血浆 Fg 水平分别为 $(3.69 \pm 0.87) \text{ g/L}$ 和 $(3.06 \pm 0.59) \text{ g/L}$ ，差异有统计学意义 ($P < 0.01$)；两组 T 基因携带者与 CC 基因型者 Fg 水平差异有统计学意义 ($P < 0.01$)；按性别分组后，病例组男性 T 基因携带者与 CC 基因型者血浆 Fg 水平差异有统计学意义 ($P < 0.01$)，对照组男性及 2 组女性 T 基因携带者血浆 Fg 水平均高于相应 CC 基因型者 ($P < 0.05$)；病例组或对照组内不同性别间 Fg 水平差异无统计学意义 ($P > 0.05$)。年龄对血浆 Fg 水平的影响：病例组青、中、老年 T 基因携带者血浆 Fg 水平均高于相应 CC 基因型者 ($P < 0.05$)；对照组仅青年组差异有统计学意义 ($P < 0.05$)；病例组不同年龄段间血浆 Fg 水平差异无统计学意义 ($P > 0.05$)；而对照组青年组血浆 Fg 水平低于中年组 ($P < 0.05$) 和老年组 ($P < 0.01$)，差异有统计学意义；中、老年组间差异无统计学意义 ($P > 0.05$)。具体数据见表 4-22 和表 4-23 (表头不规范，读者应学会修改)。

表 4-22 β -148C/T 基因多态性和血浆 Fg 水平 (g/L, χ^2)

基因型	病例组		对照组	
	男	女	男	女
CC	$3.43 \pm 0.82(55)$	$3.47 \pm 0.63(29)$	$2.96 \pm 0.58(36)$	$2.96 \pm 0.49(34)$
CC+TT	$3.98 \pm 0.88(43)^*$	$4.03 \pm 1.01(24)^{\#}$	$3.29 \pm 0.57(18)^{\#}$	$3.33 \pm 0.71(13)^{\#}$

注：* 与 CC 基因型比较， $P < 0.01$ ；# 与 CC 基因型比较， $P < 0.05$ 。

表 4-23 不同年龄段血浆 Fg 水平 (g/L, χ^2)

基 因 型	病 例 组			对 照 组		
	青 年	中 年	老 年	青 年	中 年	老 年
CC	3.41 ± 0.93	3.38 ± 0.62	3.55 ± 0.73	2.72 ± 0.44	3.01 ± 0.58	3.19 ± 0.51
CT	4.07 ± 1.16 *	3.89 ± 0.82 *	4.15 ± 0.92 *	3.06 ± 0.50 *	3.38 ± 0.61	3.68 ± 0.72
合计	3.66 ± 1.06	3.65 ± 0.77	3.77 ± 0.85	2.84 ± 0.49 [#]	3.12 ± 0.60	3.32 ± 0.60

注：* 与 CC 基因型比较, $P < 0.05$; # 与同组中、老年比较, $P < 0.05$ 。

【案例辨析】 原文是想在考虑不同因素的组合下比较血浆 Fg 水平差异是否有统计学意义, 用 t 检验进行了每两组的比较。从原文结果的叙述可以知道血浆 Fg 水平与年龄段、是否患病、基因型等因素有关, 因此探讨血浆 Fg 水平与其他因素的关系时不能仅用单因素分析的方法去考察, 而且从表中看出这些因素之间很可能不是互相独立的, 可能存在交互作用, 因此用单因素分析方法并不妥当。正确的分析方法应该是采用多重线性回归分析或多因素某种设计下定量资料方差分析。

【案例释疑】 对于本研究资料进行多重线性回归, 将“血浆 Fg 水平(将数据做对数转换)”作为因变量, “是否患病”、“性别”、“基因型”和“年龄”作为自变量进行多重线性回归。其中“是否患病”、“性别”、“基因型”为二分类变量, 赋值为“0”或“1”; “年龄”用原始值作为连续变量引入模型, 如有必要可以在自变量中添加交互作用项, 最终可以考察在平衡了其他影响因素之后某因素和血浆 Fg 水平的关系, 而且可以得到相应的回归方程。当然, 若关心的结果变量为“是否患病”, 就需要进行多重 logistic 回归分析。

4.4.2 Logistic 回归分析方面存在的问题

【案例 4-21】 原文研究目的探讨 $\beta 1$ 肾上腺素能受体基因 389A/G 多态性与急性心肌梗死 (AMI) 的关系。检测了 150 名急性心肌梗死患者和 150 例性别和年龄匹配的对照者的基因型, 同时采集相关的临床资料, 进行病例对照统计学分析, 采用 Binary 回归分析评价影响急性心肌梗死发生的危险因素, 所有检验以 $P < 0.05$ 为差异有统计学意义。在多因素参与的回归分析中, 389A/G 多态性 (OR: 2.88, 95% CI: 1.70 ~ 4.88, $P < 0.01$)、吸烟 (OR: 2.72, 95% CI: 1.52 ~ 4.88, $P < 0.01$)、高脂血症 (OR: 2.85, 95% CI: 1.68 ~ 4.86, $P < 0.01$)、糖尿病 (OR: 2.38, 95% CI: 1.27 ~ 4.47, $P < 0.01$) 和高血压 (OR: 2.00, 95% CI: 1.62 ~ 3.45, $P < 0.05$) 均为 AMI 的独立危险因素。具体见表 4-24。

表 4-24 AMI 危险因素 Binary Logistic 回归分析结果

项 目	比 值 比	95%CI	P 值
年龄	0.99	0.96 - 1.01	> 0.05
性别(男)	0.82	0.43 - 1.56	> 0.05
高血压	2.00	1.62 - 3.45	< 0.05
糖尿病	2.38	1.27 - 4.47	< 0.01
高血脂症	2.85	1.68 - 4.86	< 0.01
吸烟	2.72	1.52 - 4.88	< 0.01
基因多态性	2.88	1.70 - 4.88	< 0.01

【案例辨析】 首先原文缺乏必要的变量赋值说明, 因为结果的解释必需要知道变量是怎么赋值的。基因多态性 AA、AG、GG 在分析时必须定义其为名义变量, 故应设置两个哑变量, 每个哑变量都为二值名义变量, 只能取值 0 或 1。假定为 X_1 和 X_2 , 可以将两个哑变量的值设置为: 当 $X_1 = 0$ 且 $X_2 = 0$ 时表示 GG, 以此作为对照的基础; 当 $X_1 = 1$ 且 $X_2 = 0$ 时表示 AG, 当 $X_1 = 0$ 且 $X_2 = 1$ 时表示 AA。

原文得到的 logistic 回归分析结果是极其不规范的, 缺乏很多必要的结果, 例如回归系数估计值 β 及其标准误、Wald 值和具体的 P 值大小等。同时原文也缺乏对变量筛选方法以及引入和剔除检验水准的说明。从原文看出年龄和性别的 P 值均大于 0.05, 说明原文很可能没有进行变量筛选, 无统计学意义的因素在经过变量筛选后都不会保留在最终的 logistic 回归方程中。

【案例释疑】 首先应给出急性心肌梗死可能的影响因素及其赋值, 见表 4-25。

表 4-25 急性心肌梗死可能的影响因素与赋值

因 素	变 量 名	赋 值 说 明
年龄	X_1	原始数值
性别	X_2	女 = 0, 男 = 1
合并高血压	X_3	否 = 0, 是 = 1
合并糖尿病	X_4	否 = 0, 是 = 1
合并高血脂症	X_5	否 = 0, 是 = 1
吸烟	X_6	不吸烟 = 0, 吸烟 = 1
基因型	X_7	GG = 1, AG = 2, AA = 3(定义为名义变量, 纳入模型时由编程法或软件自动生成哑变量)
急性心肌梗死	Y	对照 = 0, 病例 = 1

正确的 logistic 回归分析结果的形式应该如表 4-26 所示。同时要写明变量筛选的方法、引入和剔除检验水准。保留在模型中的变量的 P 值都应小于剔除检验水准。

表 4-26 急性心肌梗死(AMI)影响因素的 logistic 回归分析结果

选 入 变 量	回归系数估计值 β	标准误 SE	Wald χ^2 值	P 值	OR 值	95%CI
...
...
...

【案例 4-22】 原文目的探讨 STAMP 2 基因功能区多态位点与新疆维吾尔族人原发性高血压的相关性, 采用以流行病学调查为基础的病例对照研究, 选取 2047 个维吾尔族人(包括 810 例高血压病患者和 1237 名对照)作为研究对象。高血压患病危险因素用 logistic 回归分析, 结果显示 3 个变异位点 rs8122、rs1981529、rs34741656 都不是高血压患病的危险因素($P > 0.05$)。Logistic 回归分析结果见表 4-27。

表 4-27 Logistic 回归分析

指 标	β	SE	χ^2 值	P 值	OR 值	95%CI
年龄	-0.067	0.005	192.083	0.000	0.936	0.926 - 0.945
性别	-0.045	0.113	0.159	0.690	0.956	0.766 - 1.193
体重指数	-0.141	0.012	158.078	0.000	0.868	0.848 - 0.889
吸烟	0.231	0.133	3.080	0.082	1.259	0.971 - 1.633
饮酒	-0.077	0.182	0.178	0.672	0.926	0.649 - 1.322
rs8122	0.050	0.105	0.224	0.636	1.051	0.855 - 1.291
rs1981529	-0.107	0.111	0.814	0.368	0.899	0.712 - 1.134
rs34741656	-0.071	0.108	0.437	0.508	0.931	0.754 - 1.150

【案例辨析】 首先原文缺乏必要的变量赋值说明, 因为变量的赋值不同结果的解释也不同。例如, 吸烟的回归系数估计值 β 为 0.231, 究竟它是危险因素还是保护因素, 与吸烟这个变量的赋值有关, 也与因变量 Y 的赋值有关。对于 3 个变异位点 rs8122、rs1981529、rs34741656, 原文很可能把这 3 个变量作有序变量处理。正确的处理是把这 3 个变量都定义为名义变量, 每个变量都产生两个哑

变量,每个哑变量都为二值名义变量,只能取值 0 或 1。例如,位点 rs8122,两个哑变量假定为 X_1 和 X_2 ,当 $X_1=0$ 且 $X_2=0$ 时表示 GG,以此作为对照的基础;当 $X_1=1$ 且 $X_2=0$ 时表示 GA,当 $X_1=0$ 且 $X_2=1$ 时表示 AA。其他两个位点类同。从原文看出性别、饮酒、rs8122、rs1981529、rs34741656 的 P 值均大于 0.10,说明原文没有进行变量筛选,无统计学意义的因素在经过变量筛选后都不会保留在最终的 logistic 回归方程中。

【案例释疑】 首先应给出新疆维吾尔族人原发性高血压可能的影响因素及其赋值,见表 4-28。做 logistic 回归分析要进行变量筛选,同时要给出引入和剔除检验水准,最终保留在回归方程中的变量的 P 值都应小于剔除检验水准。

表 4-28 新疆维吾尔族人原发性高血压可能的影响因素及其赋值

因 素	变 量 名	赋 值 说 明
年龄	X_1	原始数值
性别	X_2	女 = 0, 男 = 1
体重指数	X_3	原始数值
吸烟	X_4	不吸烟 = 0, 吸烟 = 1
饮酒	X_5	不饮酒 = 0, 饮酒 = 1
rs8122	X_6	GG = 1, GA = 2, AA = 3(定义为名义变量,纳入模型时自动生成哑变量)
rs1981529	X_7	AA = 1, AG = 2, GG = 3(定义为名义变量,纳入模型时自动生成哑变量)
rs34741656	X_8	GG = 1, GA = 2, AA = 3(定义为名义变量,纳入模型时自动生成哑变量)
原发性高血压	Y	对照 = 0, 病例 = 1

【案例 4-23】 原文研究载脂蛋白 A5(APOA5) - 12238T/C 多态性与新疆维吾尔族冠状动脉粥样硬化性心脏病及血脂水平的相关性。对 344 例冠心病患者和 408 例对照的 APOA5 基因 - 12238T/C 多态性进行检测。将是否患冠心病作为因变量,性别、吸烟史、饮酒史、高血压史、糖尿病史、BMI、血清 TC、TG、HDL-C、LDL-C 水平及 APOA5 基因多态性作为自变量,作多因素非条件 Logistic 回归分析,逐个选入或剔除(前进法和后退法)不显著变量, α 取 0.05。结果见表 4-29。从表 4-29 看出,入选主效应模型的变量有吸烟史、高血压、血清 TC、HDL-C 水平及基因多态性。在校正了上述因素后, - 12238CC 基因型与 TT 型比较,患冠心病的风险性明显减小(OR = 0.328, 95% CI: 0.154 ~ 0.700),并具有统计学意义($P < 0.01$);而 TC 基因型与 TT 型比较,差异无统计学意义。

表 4-29 多因素非条件 logistic 回归分析结果

危 险 因 素	β	SE	wald χ^2	P 值	OR 值	95%CI
CC 基因型	-1.114	0.386	8.315	0.004	0.328	0.154 - 0.700
高血压	0.989	0.228	18.906	<0.01	2.689	1.722 - 4.200
糖尿病	0.808	0.387	4.365	0.037	2.243	1.051 - 4.788
吸烟	1.663	0.289	33.158	<0.01	5.276	2.995 - 9.293
饮酒	-0.713	0.288	33.158	0.013	0.490	0.279 - 0.862
总胆固醇	0.673	0.118	32.728	<0.01	1.959	1.556 - 2.467
高密度脂蛋白胆固醇	-1.861	0.620	9.015	0.018	0.351	0.148 - 0.834

注: Logistic 回归模型校正年龄、性别、BMI、吸烟史、饮酒史、高血压史、糖尿病史、血清 TC、TG、HDL-C、LDL-C 水平。

【案例辨析】 原文缺乏必要的变量赋值说明,因为不同的赋值会影响结果的解释。一般情况下年龄也应该作为自变量进行考察,而原作者未将其作为自变量。

【案例释疑】 应给出新疆维吾尔族冠心病可能的影响因素及其赋值情况,见表 4-30。

表 4-30 新疆维吾尔族冠心病可能的影响因素及其赋值

因 素	变 量 名	赋 值 说 明
年龄	X_1	原始数值
性别	X_2	女 = 0, 男 = 1
吸烟史	X_3	无 = 0, 有 = 1
饮酒史	X_4	无 = 0, 有 = 1
高血压史	X_5	无 = 0, 有 = 1
糖尿病史	X_6	无 = 0, 有 = 1
BMI	X_7	原始数值
血清总胆固醇	X_8	原始数值
三酰甘油	X_9	原始数值
高密度脂蛋白胆固醇	X_{10}	原始数值
低密度脂蛋白胆固醇	X_{11}	原始数值
APOA5 基因多态性	X_{12}	TT = 1, TC = 2, CC = 3 (定义为名义变量, 纳入模型时自动生成哑变量)
冠心病	Y	对照 = 0, 病例 = 1

4.4.3 本节相关知识简要回顾与小结

由于事物间的联系是多方面的, 通常影响结果变量的因素不止一个, 因此需要采用较为复杂的多因素分析方法。多重线性回归分析就是处理因变量是连续变量的多因素分析方法, 它可以在平衡了其他影响因素的情况下考察因变量和某自变量的关系。在自变量较多的情况下, 可以根据专业知识和研究目的, 结合统计分析方法进行自变量的筛选, 常用的筛选方法有向前法、向后法和逐步回归法。使用多重线性回归时要注意自变量之间是否存在多重共线性, 同时依据专业知识考虑是否要考察变量间的交互作用等。由于多重线性回归对资料有要求, 因此建立回归方程后还需要计算残差, 检验残差是否服从正态分布。

Logistic 回归常用于因变量是分类变量的多因素分析, 是病例对照研究关联分析中最常用的方法, 它可以控制其他混杂因素的影响来看某基因多态性和疾病的关联。最常用的是因变量为二分类的 logistic 回归分析; 如果因变量是有序变量则用有序 logistic 回归分析; 如果因变量是无序多分类资料则用多分类 logistic 回归分析。使用 logistic 回归分析时要注意变量的取值形式, 变量的取值不同, 参数的含义、量值及符号都可能发生变化。如果自变量是多值名义变量, 必需要引入哑变量处理, 这样才可以对参数的实际意义作出解释, 同时可以计算名义变量各水平间相互比较的优势比 (OR)。当自变量个数较多时, 需要进行变量的筛选, 尽可能将效果显著的自变量选入模型, 将作用不显著的自变量排除。逐步回归是最常用的筛选变量的方法, 它要求设置选入自变量的检验水准 $\alpha_{\text{入}}$ 和剔除自变量的检验水准 $\alpha_{\text{出}}$ 。但也不能过分依赖软件, 实际当中最好根据专业背景和研究目的去选择变量。

参 考 文 献

- [1] 孙振球. 医学统计学 (第3版). 北京: 人民卫生出版社, 2010.
- [2] 方积乾. 卫生统计学 (第6版). 北京: 人民卫生出版社, 2008.
- [3] 胡良平. 统计学三型理论在实验设计中的应用. 北京: 人民军医出版社, 2006.
- [4] 胡良平. 生物医学科研课题的研究设计与统计分析——错误案例辨析与释疑. 北京: 军事医学科学出版社, 2008.
- [5] 胡良平, 郭晋. 医学遗传统计分析 with SAS 应用. 北京: 人民卫生出版社, 2011.
- [6] 曹瑞, 钱俊, 陈平雁. 定性数据的多重比较问题. 南方医科大学学报, 2010, 30(1): 118-120.
- [7] Li R, Xue C, Li C, et al. TRAC Variants Associate with IgA Nephropathy. J Am Soc Nephrol, 2009, 20: 1359-1367.

第 5 章 疫苗和免疫统计设计与数据分析

免疫学是研究机体免疫系统结构和功能的科学,主要研究内容包括免疫系统的组织结构、免疫系统对自身和非自身的识别及应答、免疫系统对非自身的排异效应及其机制、免疫耐受的诱导、维持、破坏及其机制等。医学免疫学还探讨免疫功能异受所致的病理过程及其机制,以及免疫学理论、方法和技术在疾病预防、诊断和治疗中的应用等。

在这一章,主要处理两个出现在免疫相关期刊中的统计学问题:计数资料的统计分析方法和 Meta 分析。本章的实际例子主要来源于疫苗和免疫期刊的例子。

下面首先强调计数资料的特点,然后介绍两种常用的基本计数资料分析模型——Poisson 回归和负二项回归。然后讨论广泛应用于循证医学中的 Meta 分析,介绍 Meta 分析的步骤,在 Meta 分析中关于固定效应(Fixed Effect)和随机效应(Random Effect)模型的争执由来已久,为了避免这个问题,本章从一种全新的角度——贝叶斯统计阐述 Meta 分析,为读者提供与传统频率统计不一样的观点。最后给出实际例子。

5.1 计数资料模型

5.1.1 计数资料的特点

计数资料(Count Data)是指单位时间、空间、面积内某个医学事件(Event)的发生次数,结果变量取值为 0、1、2、...等非负的整数,事件数表现为离散型数值。如鼻喷嚏平均次数、慢性阻塞性肺疾病急性加重发作次数和 gamma 干扰素治疗慢性肉芽肿后病人复发次数等。

计数资料有类似定量资料的性质,在很多时候被简单地当做连续资料处理,采用 t 检验或方差分析比较不同组间某个事件数的均数是否存在差别,如采用 t 检验比较西洛司特组和对照组在治疗慢性阻塞性肺疾病后急性加重发作次数的差别;采用方差分析比较不同组大鼠在艾叶挥发油治疗变应性鼻炎后鼻喷嚏平均次数的差别。或者根据没有发生和发生次数大于 0 次分类当做二分类资料进行 χ^2 检验、Logistic 回归,或者将事件数当做顺序变量的取值进行有序 Logistic 回归(Ordinal Logistic)。然而,在大部分情况下,这些方法都不适用于计数资料的分析。计数资料作为一种特殊的离散(Discrete)和受限变量(Limited Variable)数据,有其自身的特点,需要采用特殊的方法进行处理。

5.1.2 Poisson 分布与 Poisson 回归

在数理统计中,Poisson 分布有着悠久的历史,最早可追溯到 1838 年。对当时广泛研究的二项分布(Binomial Distribution),在事件的发生概率 p 很小、试验次数 n 很大的情况下,法国数学家 Poisson(1781.6.21—1840.4.25)推导出了二项分布的极限分布,为了纪念他,故将他研究出的分布称为 Poisson 分布。其早期一个著名的应用例子是 Bortkiewicz 观察到在普鲁士的骑兵部队中,每年被马踢死的士兵数服从 Poisson 分布。

设随机变量 y 服从参数为 n, p 的二项分布,记作 $y \sim \text{Bin}(n, p)$,二项分布中随机变量 y 的概率分布为:

$$P(y) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, 1, 2, \dots, n \quad (5-1)$$

$P(y)$ 表示了在 n 个独立的 Bernoulli 试验中成功次数为 y 的概率。在试验次数 n 足够大、事件 A 发生的概率 p 很小时, 事件 A 的发生次数 y 近似服从参数为 μ 的 Poisson 分布 ($\mu = np$), 即当 $n \rightarrow \infty$, $p \rightarrow 0$ 并且 $np = \mu$ 时, 二项分布的极限分布就是 Poisson 分布, 其概率密度函数为:

$$\lim_{n \rightarrow \infty, p \rightarrow 0} \left[\binom{n}{y} \left(\frac{\mu}{n} \right)^y \left(1 - \frac{\mu}{n} \right)^{n-y} \right] = \frac{\exp(-\mu) \mu^y}{y!}, y = 0, 1, 2, \dots \quad (5-2)$$

式中, $\mu = E(y)$ 表示观察时间内观察单位事件发生次数的期望值, 称为事件的平均发生率或事件的均数。此原理称为 Poisson 过程的稀有事件法则 (Law of Rare Events), 即在试验的次数足够大、事件发生的概率足够小并且满足事件发生之间相互独立, 事件数服从或近似服从 Poisson 分布, 图 5-1 展示了 Poisson 分布的概率密度函数图。

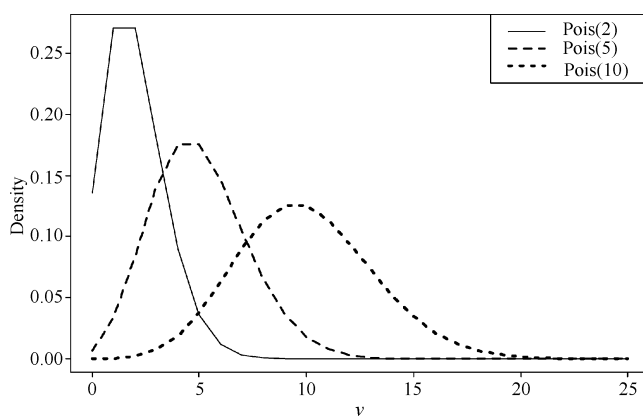


图 5-1 Poisson 分布的概率密度函数图

Poisson 分布有以下重要特征。

(1) 对于服从 Poisson 分布的 m 个相互独立的随机变量, 它们的和也服从 Poisson 分布, 均数为 m 个随机变量均数之和, 即 Poisson 分布具有可加性。

(2) 随着 μ 的增加随机变量取“0”值的概率逐渐降低, 例如均数为 1、2、3、5 和 10 时随机变量为 0 的概率分别为 36.8%、13.5%、5.0%、0.7% 和 0.00454%; 并且随着 μ 的增加 Poisson 分布的中心逐渐右移呈钟形曲线, 在 μ 大于 10 时(或更大些, 如 20)随机变量的分布近似呈正态分布。

(3) 总体均数与总体方差相等是 Poisson 分布最为熟知的特点, 即 $E(Y) = \text{Var}(Y) = \mu$, 称为等离散 (Equi-dispersion)。

Poisson 回归是计数资料分析的基本模型。当 μ 受到自变量 x 影响时, 通过指数形式的参数化与自变量 x 建立关系, 得到 Poisson 回归模型 (Poisson Regression Model, PRM), 其形式记作:

$$\mu = E(y_i | x_i) = \exp(\beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) = \exp(X'\beta) \quad (5-3)$$

式中, X 、 β 分别为自变量向量和参数向量。自变量组合以指数形式的参数化就满足了 μ 在实际应用中必须大于 0 的限制条件, 文献中将式 (5-3) 称为指数均数函数 (Exponential Mean Function), 对式 (5-3) 两边取对数:

$$\log(\mu) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p = X'\beta \quad (5-4)$$

这样就将指数形式转化为线性表达式, 因此 Poisson 回归又称为对数线性模型 (Log-linear Models)。

Poisson 回归在应用中需要满足以下两个条件: 事件的发生相互独立; 事件的条件均值等于条件方差。

5.1.3 负二项分布与负二项回归

Poisson 分布要求事件发生前后之间相互独立并且事件的均数和方差相等,而实际中的计数资料常常不满足这些条件并且要更为复杂,如发生事件之间的非独立性(如具有传染性、遗传性、季节波动性、地方性和一些原因不明的聚集现象等),以及事件的均数和方差悬殊(主要是方差大于均数)的情况,此时即使均数结构指定正确,Poisson 回归也是有问题的。Greenwood 和 Yule 以及 Eggenberger 和 Polya 在 Poisson 分布的基础上扩展得到了负二项分布。

负二项分布的概率密度函数:

$$P(y) = \frac{\Gamma(y + \alpha^{-1})}{y! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^y \quad (5-5)$$

α 为离散或尺度参数 (Dispersion/Scale Parameter), 反映发生次数 y 的变异程度。当 α 等于 0 时, 负二项分布便退化为 Poisson 分布, α 越大, 表示事件数的变异越大。

负二项分布的均数 $E(y) = \mu$, 但方差 $\text{Var}(y) = \mu(1 + \alpha\mu) = \mu + \alpha\mu^2$, 可见, 在 α 大于 0 时, 负二项分布的方差大于均数, 能够容纳比 Poisson 分布更大的变异。

负二项分布的概率密度函数还可写为如下的形式:

$$P(y) = \frac{\Gamma(y + n)}{y! \Gamma(n)} (p)^n (1 - p)^y \quad (5-6)$$

式(5-6)和式(5-5)的关系很明显。图 5-2 展示了负二项分布的概率密度函数图。

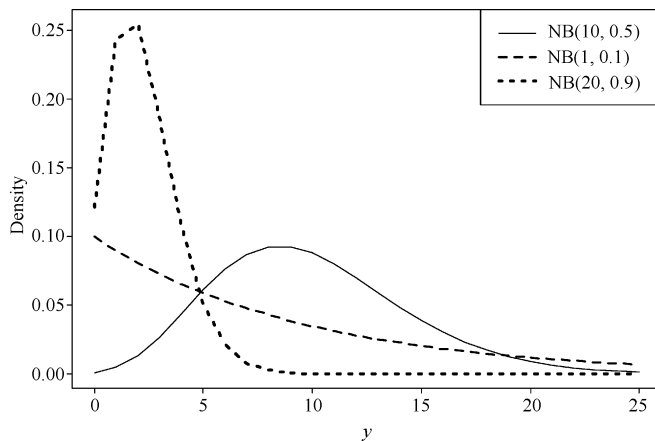


图 5-2 根据式(5-6)显示的负二项分布概率密度函数图

按照 Poisson 回归参数化方法将 μ 与自变量 x 联系起来, 得到负二项回归 (Negative Binomial Regression Model, NBRM):

$$\mu = \exp(\beta_0 + x_1\beta_1 + \cdots + x_p\beta_p) = \exp(X'\beta) \quad (5-7)$$

5.1.4 过度离散检验

事件的条件方差超过条件均数时称为过度离散 (Over-Dispersion)。在包含自变量后在一定程度上会减小事件数的变异, 然而在横断面资料中, 自变量所能解释的数据变异常常不到 50%, 因此对于样本均数和样本方差相差两倍以上计数资料而言, 即使在考虑了自变量效应的情况下, 仍然可能存在过度离散。对过度离散的资料, Poisson 回归常常低估参数估计值的标准误差。过度离散检验 (Over-Dispersion Test) 可用来判断事件数的发生是否过度离散, 有基于残差和样本均数方差等多种不同的检验方法, 本章主要介绍其中几种方法。

1. O 检验

O 统计量的计算为:

$$O = \sqrt{(n-1)/2} (S^2 - \bar{x}) / \bar{x} \quad (5-8)$$

式中, n 为总观察单位数, S^2 、 \bar{x} 分别为事件数的方差和均数, 在均数和方差相等的条件下统计量 O 近似服从标准正态分布。

2. 基于 Poisson 的回归残差检验

在满足 Poisson 回归的前提下, 构造回归方程:

$$(y - \hat{\mu})^2 = \lambda \hat{\mu} + \varepsilon \quad (5-9)$$

$\hat{\mu} = \exp(x'\hat{\beta})$, 为事件数的预测值。上面的检验方法即是对因变量 $(\hat{\mu})^2$ 建立自变量的不包含常数项的最小二乘回归方程, 在均数和方差相等的条件下, 对 $\lambda = 1$ 的假设检验等价于过度离散检验, 其大小反映了过度离散的程度。

或者构造另外一个回归方程, 如下:

$$(y - \hat{\mu})^2 - \hat{\mu} = \lambda \hat{\mu}^2 + \varepsilon \quad (5-10)$$

即对因变量 $(y - \hat{\mu})^2 - \hat{\mu}$ 建立自变量 $\hat{\mu}^2$ 的不包含常数项的最小二乘回归方程, 在均数和方差相等的条件下, 对 $\lambda = 0$ 的检验等价于过度离散检验。

3. 得分检验

在 Poisson 回归的前提下, 构造得分检验 (Score Test) 的统计量 T 为:

$$T = \sum_{i=1}^N [(y_i - \hat{\mu}_i)^2 - y_i] / \sqrt{2 \sum_{i=1}^N \hat{\mu}_i^2} \quad (5-11)$$

在均数和方差相等的条件下 T 统计量服从标准正态分布, N 为样本量。

4. 拉格朗日乘数检验

Greene (2002) 介绍了一种基于负二项模型的过度离散检验方法, 称为拉格朗日乘数检验。由于 Poisson 回归可以通过约束负二项模型的参数而得到, 所以能够建立拉格朗日乘数检验。统计量为

$$LM = \left[\frac{\sum_{i=1}^N w_i [(y_i - \hat{\mu}_i)^2 - y_i]}{\sqrt{2 \sum_{i=1}^N w_i \hat{\mu}_i}} \right]^2 \quad (5-12)$$

此处的权重 w_i 取决于所假定的另一种分布, 在负二项分布中, $w_i = 1$, 因此当假定 Poisson 回归通过约束负二项模型的参数而得到时, 拉格朗日乘数检验统计量又可以表示为:

$$LM = \frac{(e'e - n\bar{y})^2}{2\hat{\mu}'\hat{\mu}} \quad (5-13)$$

式中, $e = y - \hat{\mu}$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ 。在均数和方差相等的条件下, LM 统计量的极限分布是自由度为 1 的 χ^2 分布。

当 $\alpha = 0$ 时负二项分布便退化为 Poisson 分布, 因此 Poisson 和负二项回归属于嵌套模型 (Nested Model), 似然比检验:

$$LR = -2 * (\ln L_{P_{10}} - \ln L_{NB}) \sim \chi^2(1) \quad (5-14)$$

可以用来检验 $H_0: \alpha = 0$, $\ln(\cdot)$ 为对数似然值, 如果拒绝了无效假设 H_0 , 那么有理由认为负二项回归优于 Poisson 回归。

5.2 Meta 分析

5.2.1 Meta 分析的概念

全世界范围内的学者可能为了一个共同的目的而从事类似的医学研究,如临床试验中阿司匹林对心肌梗死的预防效果研究、流行病学中吸烟和肺癌关系的病例对照研究等。这些研究在设计的诸多方面如研究人群、干预措施、样本量和效应指标等存在差异,使得研究的结论不一,有的研究显示阳性结果,有的显示弱的但没有统计学意义的结果,有的甚至得出完全相反的结论。Meta 分析为这种结果不一致的类似研究提供了一种综合评价的方法,在循证医学中得到了越来越多的应用。

Meta 分析最早源于 Fisher 1920 年提出的“合并 P 值”的思想,1955 年 Beecher 首次给出初步的概念,1976 年心理学家 Glass 将通过合并统计量对文献进行综合研究的方法命名为“Meta Analysis”。

广义的 Meta 分析是对具有共同研究目的的多个研究进行综合分析的一系列过程,包括提出研究问题、检索相关文献、制定文献的纳入和排除标准、汇总基本信息、综合分析并报告结果等。狭义的 Meta 分析仅是一种单纯的合并统计量的统计学方法。

Meta 分析能够提高检验效能,解决若干单个研究结论不一致问题,改善效应量的估计,提出新的研究问题和研究方向。

5.2.2 Meta 分析的基本步骤

Meta 分析一般包括以下过程。

1. 明确研究目的,制订研究计划

同其他研究方法一样,明确研究目的是 Meta 分析中非常关键和重要的一步。研究目的通常表现为问题的形式,这些问题一般是临床研究或流行病学研究中不确定或有争议的问题。根据研究目的,制订研究计划和方案。

2. 建立检索策略,收集所有相关的研究文献和资料

检索策略是指明确检索的目的和信息需求,选择检索系统,确定检索词,构造检索式,制订出完善的检索计划和方案。常用的检索系统有 Cochrane Library、arxiv、PubMed、Medline、Springer、CBMdisc、万方、维普、CNKI,等。除计算机检索外,还需要查找现刊目录、综述性文献以及临床试验登记资料。文献的检索要全面完整,不能遗漏对分析产生重大影响的文献,包括公开发表与未发表的文献。

3. 制定文献的纳入和排除标准

制定文献纳入和排除标准即确定 Meta 分析的文献筛选条件,应考虑研究对象的确定、试验设计方法、干预因素、研究的结局变量或效应变量、样本含量、观察时间或随访年限、公开发表的时间和语种等。

4. 评价纳入文献的质量

不同类型研究的评价方法虽各不相同,如病因研究的评价、诊断性试验研究的评价、预后研究的评价等。

5. 提取必要的数据信息

从纳入的文献中提取数据信息,包括基本信息、研究特征、结果测量等内容,确定和选择要进行分析和评价的效应变量。

6. 资料的统计学分析

统计学分析是 Meta 分析最重要的步骤之一,也是 Meta 分析区别于传统文献综述的主要特征。统计分析的主要过程包括研究类型的确定、效应指标的选择、纳入研究的同质性检验、模型选择与分析、效应合并值的估计与假设检验、效应合并值的图示等。

7. 敏感性分析

敏感性分析是评价 Meta 分析结果稳定性和可靠性的常用方法,即在排除可能是异常结果的研究后,重新进行 Meta 分析,其结果与未排除时的 Meta 分析结果相比较,若敏感性分析未从实质上改变结果,表明该分析结果较为可靠;反之,说明结果的稳定性较差,解释结果和下结论时应非常慎重。

8. 报告结果

Meta 分析完成后,不仅报告统计分析结果,而且要报告整个过程的全部内容,主要有课题研究背景和对象、资料的检索方法、统计分析方法、结果和讨论等若干部分。

5.2.3 固定效应和随机效应模型

假设纳入 Meta 分析中的试验研究存在两个极端:①这些研究只不过是同一个方案的相同重复而已;②这些研究存在着相当大的差异以致不能从一个研究中得到关于另一个研究的任何信息。一个基本的原则是,如果各个纳入 Meta 分析的研究同质,则可将不同的研究结果合并采用固定效应模型(Fixed Effect Models),如 Peto 法,计算研究的总体合并效应;如果纳入 Meta 分析的研究具有不可忽略的异质性,则采用随机效应模型(Random Effect Models)或混合效应模型(Mixed Effect Models),如 DerSimonian-Laird 法,允许不同研究之间存在变异。很显然,实际应用中纳入 Meta 分析的各研究之间关系处于两者之间,它们都是为了同一个共同的医学问题而进行的,而且这些研究的确是存在差别的,但是这种差别还不足以使人相信其中的一个或某几个研究的结果。

Meta 分析的一个困难之处就在于对试验是否同质的界定,这也是 Meta 分析中最具有争论问题之一,研究显示常用的异质性 χ^2 检验效能不高,其他异质性检验的功效也有待进一步研究。

5.2.4 贝叶斯 Meta 分析

本节主要考虑在贝叶斯统计框架下进行 Meta 分析,随着方法的展开,将看到贝叶斯 Meta 分析将固定效应模型和随机效应模型同时考虑在内,很自然地解决了这个问题。

1. 临床试验和指标选择

这里主要考虑定性资料的情况。

单个定性临床试验资料多表示为四格表的形式,假设试验组中 n_1 个受试对象中有 x_1 个被观察到结果,对照组中 n_0 个受试对象中有 x_0 个被观察到结果。试验组和对照组的抽样分布可假设为两个独立的二项分布:

$$x_1 \sim \text{binary}(n_1, \pi_1), x_0 \sim \text{binary}(n_0, \pi_0)$$

Meta 分析中所关注的效应指标包括率差 $\pi_1 - \pi_0$,相对比 π_1/π_0 和优势比(Odds Ratio, OR)。由于优势比为大多数人所熟悉,并且在临床试验和病例对照研究中都具有良好的可解释性,因此主要专注于优势比的推断:

$$\text{OR} = (\pi_1 - \pi_0\pi_1)/(\pi_0 - \pi_0\pi_1) \quad (5-15)$$

设 $E = \ln(\text{OR})$,可通过样本估计:

$$y = \hat{E} = \ln(\widehat{OR}) = \ln[x_1/(n_1 - x_1)] - \ln[x_0/(n_0 - x_0)] \quad (5-16)$$

称为 OR 的 logit 变换, y 渐近服从正态分布, 其渐近方差为:

$$\sigma^2 = \text{Var}(y) = 1/x_1 + 1/(n_1 - x_1) + 1/x_0 + 1/(n_0 - x_0) \quad (5-17)$$

在样本量比较少时, 常常对每个格子增加一个较小的数如 0.5 以提高正态近似效果。

2. 等级贝叶斯模型

对 OR 作 logit 变换后, 可认为纳入 Meta 分析的 k 个研究效应 y_i , $i = 1, \dots, k$ 来自正态分布, y_i 为对应研究潜在效应 θ_i 的一个点估计值。首先建立 Meta 分析贝叶斯分层模型 (Bayesian Hierarchical Models) 的第一个层次:

$$y_i | \theta_i, \sigma_i^2 \sim N(\theta_i, \sigma_i^2) \quad (5-18)$$

σ_i^2 为对应的方差, 并假设已知可由式 (5-17) 估计得到。在 Meta 分析中, 纳入的研究样本量一般都不会太小, 因此 σ_i^2 已知的假定并不会对结果分析产生实质的影响。通过式 (5-18) 得到实际观察效应 $Y = y_1, \dots, y_k$ 的似然函数 (Likelihood Function):

$$p(Y | \theta_i, \sigma_i^2) = \prod N(\theta_i, \sigma_i^2)$$

对 θ_i 建立可互换 (Exchangeable) 的正态先验 (Prior Distribution), 得到第二个层次:

$$\theta_i | \mu, \nu^2 \sim N(\mu, \nu^2) \quad (5-19)$$

μ, ν^2 称为超参数 (Hyper-Parameters), 需要进一步指定先验分布。 θ_i 的可互换性表示在已知 μ, ν^2 的情况下, 任意置换下标 $1, \dots, k$, θ_i 的联合密度 (Joint Distribution) $p(\theta_1, \dots, \theta_k | \mu, \nu^2)$ 保持不变。这在绝大部分 Meta 分析中无疑是成立的, 因为对于纳入 Meta 分析的各研究, 除了第一个等级中处理组和对照组的样本量可作为分辨不同研究的标识外, 研究者在 Meta 分析贝叶斯分层模型的第二个等级已经不能再区别各个 θ_i 了。式 (5-19) 暗示各个研究效应 θ_i 关于 μ, ν^2 是条件独立的, $\theta = \theta_1, \dots, \theta_k$ 可看作是来自正态总体 $N(\mu, \nu^2)$ 的一个随机样本, 或者说效应服从一个共同的总体分布。 μ 称为 Meta 分析中研究的真实效应, 是 Meta 分析中统计推断的目标, ν^2 测量了纳入 Meta 分析的各研究的同质性程度, 当 $\nu^2 \rightarrow 0$, 潜在效应 θ_i 之间没有变异, 意味着各研究是彼此的重复, 对应 Meta 分析的固定效应模型; 随着 ν^2 增大, 表示潜在效应 θ_i 存在变异, 也即是各研究具有随机效应; $\nu^2 \rightarrow \infty$ 意味着各研究彼此互不相关, 对应着前文中第二种极端状况。对 μ 指定关于 ν^2 无信息 (Non Informative) 的均匀先验, $p(\mu | \nu^2) \propto 1$, 此时 μ, ν^2 的先验分布为 $p(\mu, \nu^2) = p(\mu | \nu^2)p(\nu^2) \propto p(\nu^2)$, $p(\nu^2)$ 为超参数 ν^2 的先验分布。

3. 联合后验分布

Meta 分析中各研究效应 y_i , σ_i^2 为实际观察到的数据, θ_i 和 μ, ν^2 都为未知参数, 根据贝叶斯定理, 各研究效应 θ_i 和超参数 μ, ν^2 的联合后验分布 (Posterior Distribution) 为:

$$\begin{aligned} p(\theta, \mu, \nu^2 | Y, \sigma_i^2) &\propto p(\mu, \nu^2) \prod N(y_i | \theta_i, \sigma_i^2) N(\theta_i | \mu, \nu^2) \\ &\propto p(\nu^2) \prod N(y_i | \theta_i, \sigma_i^2) N(\theta_i | \mu, \nu^2) \end{aligned} \quad (5-20)$$

σ_i^2 假设已知并由式 (5-17) 估计。在大多数的 Meta 分析中, 式 (5-20) 都比较复杂, 后验推断将难以进行。通过简化得到 θ_i 关于超参数 μ, ν^2 的条件后验 (Conditional Distribution) 分布为:

$$\begin{aligned} \theta_i | Y, \mu, \nu^2, \sigma_i^2 &\sim N(\mu_i, \tau_i) \\ \mu_i &= \frac{y_i/\sigma_i^2 + \mu/\nu^2}{1/\sigma_i^2 + 1/\nu^2}, \quad \tau_i^{-1} = 1/\sigma_i^2 + 1/\nu^2 \end{aligned} \quad (5-21)$$

μ 关于 ν^2 的条件后验分布为:

$$\begin{aligned}\mu | Y, \nu^2, \sigma_i^2 &\sim N(\mu_0, V) \\ \mu_0 &= \frac{\sum y_i / (\sigma_i^2 + \nu^2)}{\sum 1 / (\sigma_i^2 + \nu^2)}, V^{-1} = \sum 1 / (\sigma_i^2 + \nu^2)\end{aligned}\quad (5-22)$$

ν^2 的后验分布为: $p(\nu^2 | Y, \sigma_i^2) \propto p(\nu^2) p(Y | \nu^2, \sigma_i^2)$ 。其中:

$$p(Y | \nu^2, \sigma_i^2) \propto V^{1/2} \prod (\sigma_i^2 + \nu^2)^{-1/2} \exp\left[-\frac{(y_i - \mu_0)^2}{2(\sigma_i^2 + \nu^2)}\right] \quad (5-23)$$

4. 经验贝叶斯 Meta 分析

所谓经验贝叶斯 (Empirical Bayes) 分析就是直接应用实际观察数据估计先验分布中的超参数, 然后用超参数的估计值进行后验分析。式(5-23)描述了在参数 ν^2 条件下实际观察效应 Y 的似然函数, 因此可对式(5-23)采用极大似然法得到参数 ν^2 的极大似然估计值 $\hat{\mu}_{ml}^2$, 然后用 $\hat{\mu}_{ml}^2$ 取代上述各式中的 ν^2 , 从而对 θ 和 μ 进行分析。其他方法如限制性极大似然估计 (Restricted Maximum Likelihood) 和一阶矩估计 (Method of Moment) 也可用来估计 ν^2 。

5. 完全贝叶斯 Meta 分析

为了完成 Meta 分析的完全贝叶斯分层模型, 还必须为参数 ν^2 指定先验分布, 为了简单给 ν^2 指定无信息先验 $p(\nu^2) \propto 1$ 或者弱信息 (Low Informative) 先验 $\text{Ga}(10^{-3}, 10^{-3})$, Ga 表示 gamma 分布。需要注意的是当给方差 ν^2 指定非正则 (non-Proper) 先验时, 需要检验导致的后验是否为正规的密度。对于分层模型中的方差成分 ν^2 而言, 通常看似正确的无信息先验 $p[\log(\nu^2)] \propto 1$ 会导致非正规后验密度, 而 $p(\nu^2) \propto 1$ 却不会。

根据前面各个参数的条件和边际后验分布 (Marginal Posterior Distribution), 可将联合后验分布分解为条件概率的形式, 这种分解有利于后验推断的 Monte Carlo 模拟抽样。

$$\begin{aligned}p(\theta, \mu, \nu^2 | Y, \sigma_i^2) &= p(\nu^2 | Y, \sigma_i^2) p(\mu | Y, \nu^2, \sigma_i^2) \\ p(\theta | Y, \mu, \nu^2, \sigma_i^2)\end{aligned}\quad (5-24)$$

对于上述描述的 Meta 分析贝叶斯分层模型而言, 通过基于模拟抽样的参数后验分布计算将变得十分容易。后验模拟抽样的策略如下。

(1) 通过非迭代的如网格点抽样 (Grid of Points Sampling)、拒绝 - 接受抽样 (Acceptance-Rejection Sampling)、重要抽样 (Important Sampling) 等, 或迭代的 Monte Carlo 模拟如 Gibbs 抽样或 Metropolis-Hastings 法等, 从 $p(\nu^2 | Y, \sigma_i^2)$ 抽样得到 ν^2 的随机样本 ν_i^2 。

(2) 以 ν_i^2 为条件从 $p(\mu | Y, \nu_i^2, \sigma_i^2)$ 中抽取 μ 的随机样本 μ_i 。

(3) 从 $\theta_i | Y, \mu_i, \nu_i^2, \sigma_i^2 \sim N(\mu_i, \tau_i)$ 中抽取潜在效应 θ_i 的随机样本 θ_{ii} 。

由于后两步的抽样都只涉及常规的正态分布, 因此抽样十分简单并且高效。

6. 网格抽样

按照式(5-24)建立的后验模拟抽样策略, 采用网格抽样法抽取 ν^2 , 设 ν^2 的先验为 $p(\nu^2) \propto 1$ 。由于 ν^2 不能为 0, 因此从一个极小的数值开始抽样, 本章中 ν^2 选取区间 (0.001, 5), 网格值的步长为 0.0001, 即 ν^2 依次取值为 0.0011, 0.0012, 0.0013, ..., 5, 则 ν^2 的取值个数为 49991。 ν^2 的网格抽样具体如下。

(1) 对每一个 ν_i^2 按式(5-25)计算其相对密度 f_i , 然后汇总计算每个 ν^2 的概率 $p_i = f_i / \text{sum}(f_i)$;

$$p(\nu^2) V^{1/2} \prod (\sigma_i^2 + \nu^2)^{-1/2} \exp\left[-\frac{(y_i - \mu_0)^2}{2(\sigma_i^2 + \nu^2)}\right] \quad (5-25)$$

(2)对全部 49991 个 ν_i^2 按照抽样概率为 p_i 进行 10^4 次有放回的抽样。

根据每一个 ν^2 的后验样本 $\nu_i^2, t = 1, \cdots, 10^4$ 从 μ 的条件后验分布 $N(\mu_0, V)$ 中抽取真实效应 μ 的后验样本 $\mu_i, t = 1, \cdots, 10^4$; 然后以 ν_i^2, μ_i 为条件从式(5-21)中抽取各研究潜在效应 θ_i 的后验样本 $\theta_{it}, i = 1, \cdots, 14, t = 1, \cdots, 10^4$ 。

对经验贝叶斯而言, 只需用 ν^2 极大似然估计值 $m\ell^2$ 取代式(5-22)中的 ν^2 , 然后从对应的后验分布抽取 μ_i^{EB} , 以 μ_i^{EB} 为条件抽取潜在效应 θ_i 的后验样本 θ_{it}^{EB} , EB 表示后验样本是以经验贝叶斯的方法得到的。

5.3 实例分析

在这一节中, 通过实际例子介绍计数资料的统计分析和 Meta 分析。前文中《艾叶挥发油治疗大鼠变应性鼻炎的试验研究》、《西洛司特对 COPD 患者炎症水平的调节作用》分析过计数资料, 同样的情况也出现在文献《长期应用糖皮质激素治疗对 RRMS 患者临床复发和神经功能残疾进展的影响》、《莲房原花青素对东莨菪碱所致小鼠记忆获得性障碍的改善作用》、《侧脑室注射链脲酶素导致大鼠脑内能量代谢障碍和视空间记忆障碍》、《 ω -3 多不饱和脂肪酸对老年痴呆大鼠学习记忆和海马 MDA、SOD 的影响》和《TNF- α 早期表达在大鼠急性心肌梗死室性心律失常发生中的作用及机制研究》。

Meta 分析也广泛地应用于免疫学中, 见文献《干扰素调节因子 5 基因 rs2004640 多态性与系统性红斑狼疮遗传易感性的荟萃分析》、《中国多发性硬化患者肿瘤坏死因子- 基因多态性的 Meta 分析》、《口服轮状病毒减毒活疫苗保护效果研究 Meta 分析》、《国内流行性感胃灭活疫苗的保护效果研究 Meta 分析》和《Oka 株水痘减毒活疫苗流行病学保护效果 Meta 分析》。在这些文献中, 大多作者基本上都很少报告原始数据。

由于不能得到论文的实际数据, 下面通过另外的数据展示 Poisson 回归、负二项回归和 Meta 分析。

5.3.1 计数资料分析

1. 实例

某临床医师收集了 128 例慢性肉芽肿患者, 随机分为两组后分别给予 gamma 干扰素和安慰剂两种治疗方法, 两组人数分别为 63 人和 65 人, 以此评价 gamma 干扰素治疗慢性肉芽肿的疗效, 评价指标为接受不同治疗方法的患者在 1 年内慢性肉芽肿复发的次数。复发次数和患者资料一般描述见表 5-1 和表 5-2, 采用 gamma 干扰素治疗者 1 年内复发的平均次数为 0.32 次, 采用安慰剂治疗者为 0.86 次, 平均复发次数为 0.59 次, 方差为 1.20, 复发次数的方差约为均数 2 倍。

表 5-1 慢性肉芽肿病复发的次数

复 发 次 数	例 数	频 率	累 计 频 率
0	84	65.63	65.63
1	27	21.09	86.72
2	9	7.03	93.75
3	5	3.91	97.66
4	1	0.78	98.44
5	1	0.78	99.22
6	0	0.00	99.22
7	1	0.78	100.00

表 5-2 患者资料描述和赋值说明

变 量	变量标签	均 数	标 准 差	最 小 值	最 大 值	赋 值 说 明
treat	处理因素	1.508	0.502	1	2	1 = 干扰素 2 = 安慰剂
age	年龄	14.641	9.844	1	44	岁
use1	使用皮质醇	1.977	0.152	1	2	1 = 是 2 = 否
use2	使用抗生素	1.133	0.341	1	2	1 = 是 2 = 否
sex	性别	1.188	0.392	1	2	1 = 男 2 = 女
count	复发次数	0.594	1.097	0	7	次数

本资料中, $n = 128$, $\bar{x} = 0.59$, $S^2 = 1.20$, 经 O 检验得 O 统计量为 8.24, 显然表示复发次数存在过度离散, 其他检验得到类似结果。从复发次数频率及 Poisson 和负二项分布概率密度曲线图可见图 5-3, 复发次数呈偏态分布, Poisson 和负二项分布在复发 2 次及其以上都和观察到的频率相近, 复发 0 次和 1 次时 Poisson 明显低估了事件的发生概率, 而负二项分布的预测概率很接近观察的频率。

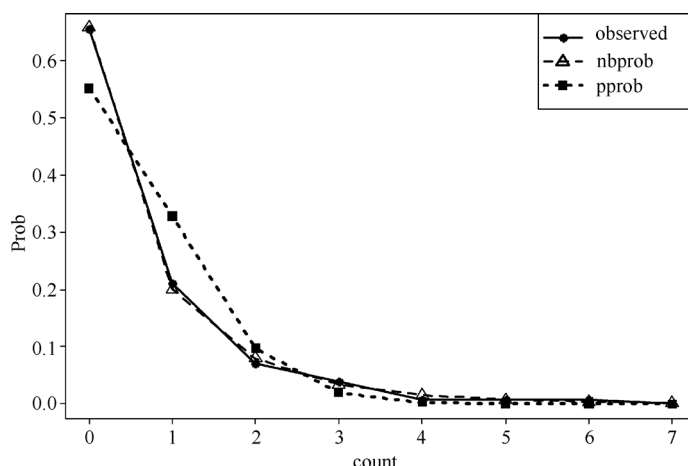


图 5-3 复发次数频率及 Poisson 和负二项分布概率密度曲线图

观察到的频率和负二项分布概率很接近, 但和 Poisson 回归相差较远。

2. 负二项回归结果

基于以上的原因, 对慢性肉芽肿复发次数的数据拟合负二项回归模型, 参数估计见表 5-3。

表 5-3 gamma 干扰素治疗慢性肉芽肿病数据的负二项回归参数估计

变 量	参数估计值	标 准 误 差	z 值	P 值	参数估计值	95% 置信区间
处理因素	0.967	0.319	3.03	0.002	0.341	1.593
年龄	-0.036	0.018	-2.01	0.044	-0.070	-0.001
使用皮质醇	-1.067	0.865	-1.23	0.217	-2.764	0.629
使用抗生素	0.471	0.433	1.09	0.277	-0.378	1.319
性别	-0.424	0.461	-0.92	0.357	-1.328	0.479
复发次数	0.424	2.178	0.19	0.846	-3.844	4.692
lnalpha	-0.035	0.436			-0.890	0.819
alpha	0.965	0.421			0.411	2.269

Likelihood-ratio test of $\alpha = 0$: $\text{chibar2}(01) = 14.21$, $\text{Prob} > = \text{chibar2} = 0.000$. 该结果通过 stata10.0 得到, 本文后面给出了对应的 SAS 程序。

Poisson 和负二项回归属于嵌套模型, 似然比检验显示 $\chi^2 = 14.21$, $P > 0$, 估计值为 0.965。负二项回归参数估计表明 gamma 干扰素治疗、患者的年龄是影响慢性肉芽肿复发次数多少的因素, 使用 gamma 干扰素者复发次数平均比使用安慰剂者少 $e^{0.967} = 2.63$ 次 ($z = 3.03$, $P = 0.002$); 患者年龄越大复发次数越少, 年龄每增加一岁复发次数平均减少 $e^{-0.036} = 0.96$ 次 ($z = -2.01$, $P = 0.044$)。尚不能认为试验前使用皮质醇、使用抗生素及患者的性别会影响复发次数。

5.3.2 贝叶斯 Meta 分析

1. 实例

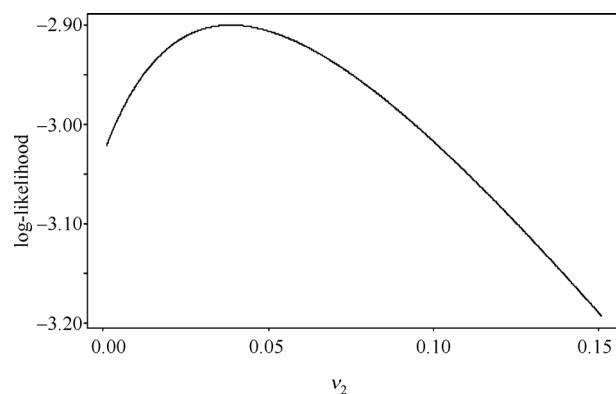
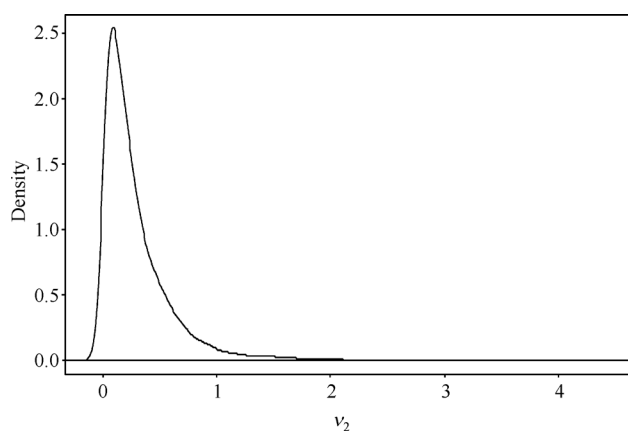
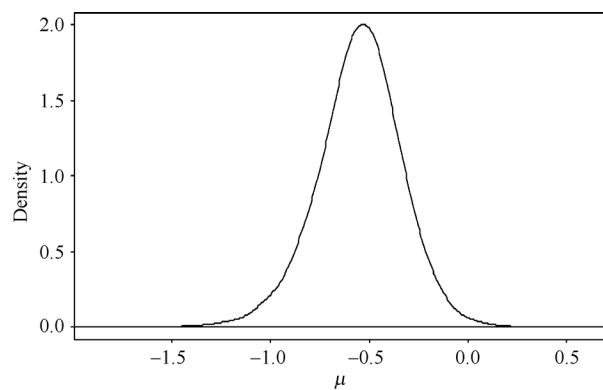
表 5-4 给出了 14 个激素预防新生儿肺透明膜病 (Hyaline Membrane Disease, HMD) 的临床试验研究数据, 激素组和对照组分别用 1 和 0 表示。 n 为每个研究中激素组或对照组的新生儿人数, x 为发生肺透明膜病人数。14 个研究中第 5 和 10 号研究显示使用激素的新生儿肺透明膜病发病率略高于对照组, 分别高出约 1.7% 和 3.3%, 第 9 和 12 号研究显示两组发病相近, 除此之外的其他研究显示使用激素的新生儿肺透明膜病发病率都低于对照组, 降低的程度在 2.3% ~ 17.3% 之间, 其中 1、6、8 号研究经四格表 χ^2 检验有统计学意义。OR 值的 logit 变换和对应的渐近方差见表 5-4 的 4 ~ 5 列。

表 5-4 14 个激素预防新生儿肺透明膜病 (HMD) 的临床试验数据及其 OR 值的 logit 变换和渐进方差以及每个研究效应的经验贝叶斯和完全贝叶斯估计值

No	激素组 (x_1/n_1)	对照组 (x_0/n_0)	y	Var(y)	y	Var	y	Var
			原始估计		经验贝叶斯		完全贝叶斯	
1	35/533	60/536	-0.584	0.049	-0.531	0.028	-0.564	0.039
2	1/61	5/54	-1.812	1.237	-0.528	0.057	-0.751	0.269
3	16/133	23/140	-0.363	0.123	-0.461	0.040	-0.431	0.076
4	3/67	7/60	-1.036	0.511	-0.530	0.053	-0.683	0.181
5	9/57	10/71	0.134	0.248	-0.407	0.048	-0.242	0.136
6	3/64	13/59	-1.749	0.448	-0.591	0.056	-0.922	0.230
7	1/71	7/75	-1.975	1.172	-0.538	0.056	-0.792	0.276
8	4/81	11/63	-1.404	0.373	-0.576	0.052	-0.842	0.187
9	32/372	33/371	-0.037	0.067	-0.325	0.033	-0.197	0.058
10	5/49	2/29	0.428	0.760	-0.445	0.056	-0.318	0.216
11	7/121	13/124	-0.646	0.238	-0.511	0.049	-0.588	0.121
12	1/24	1/22	-0.091	2.091	-0.481	0.059	-0.498	0.261
13	5/36	6/37	-0.182	0.431	-0.464	0.053	-0.417	0.156
14	6/96	8/92	-0.357	0.315	-0.472	0.051	-0.462	0.134

2. 贝叶斯 Meta 分析结果

ν^2 和 Y 的对数边际似然关系见图 5-4, HMD 数据 ν^2 的极大似然估计值为 0.039。 ν^2 基于网格抽样的后验密度见图 5-5, μ 的后验密度估计见图 5-6。真实效应 μ 的经验贝叶斯和完全贝叶斯后验结果见表 5-5, 各潜在效应 θ_i 的经验贝叶斯和完全贝叶斯后验均值和方差见表 5-4, 对应的 95% 置信区间见图 5-7。

图 5-4 真实效应的先验 ν^2 和实际效应 Y 的对数边际似然函数关系图 5-5 ν^2 基于网格抽样的后验密度函数图图 5-6 真实效应 μ 的后验密度估计表 5-5 真实效应 μ 的经验贝叶斯和完全贝叶斯后验估计以及其 OR 形式的 95% 置信区间

	经验贝叶斯				完全贝叶斯			
	均 数	方 差	下 限	上 限	均 数	方 差	下 限	上 限
μ	-0.490	0.021	0.461	0.814	-0.551	0.046	0.379	0.878

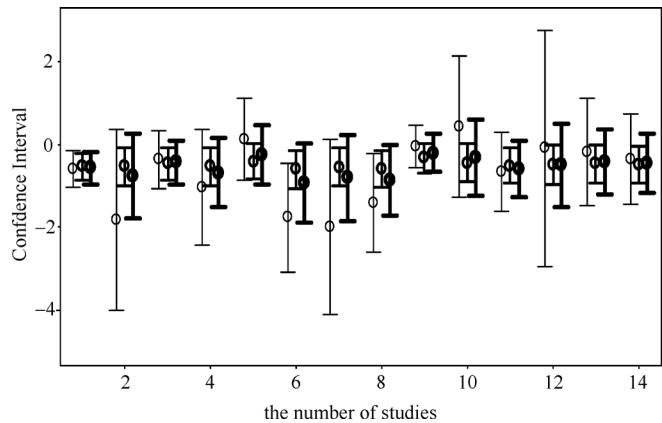


图 5-7 不同研究 OR 值 logit 转换的 95% 置信区间，从左到右依次为原始数据和经验贝叶斯及完全贝叶斯的置信区间

从图 5-6 可见真实效应 μ 的取值基本小于 0，表明在临床中使用激素会降低新生儿肺透明膜病的发生，经验贝叶斯显示相对于对照组，使用激素会使新生儿肺透明膜病大约降低 $39\% = 1 - e^{-0.490}$ ，完全贝叶斯结果显示大约降低 $42\% = 1 - e^{-0.551}$ ，两者的结果相差不大。虽然单独的 14 个试验研究显示使用激素的效果不明显，但是 Meta 分析的综合效应却是有统计学意义的，平均而言使用激素会使新生儿肺透明膜病降低约 40%。

3. 说明和注释

本章对 Meta 分析采用了贝叶斯分层模型以判别真实的综合效应，贝叶斯分层模型将纳入 Meta 分析的不同研究当作来自共同总体分布的随机样本，类似于 DerSimonian 和 Laird 的随机效应模型，贝叶斯 Meta 分析将参数看作随机变量，避免了固定效应和随机效应问题，获得的是关于参数的后验分布，也就自然地解决了频率统计估计 ν^2 的难题。分析中同时考虑了经验贝叶斯和完全贝叶斯，经验贝叶斯无须从 ν^2 的后验分布抽样，后验计算相对完全贝叶斯要简单得多，但主要不足之处在于经验贝叶斯没能考虑到先验方差 ν^2 本身所具有的不确定，从而低估了真实效应 μ 的变异。从 HMD 数据 Meta 分析后验结果可见，由于不能分析 ν^2 本身所具有的变异，所有潜在效应 θ_i 的 95% 置信区间都要比完全贝叶斯的置信区间窄，真实效应 μ 的经验贝叶斯后验方差也小于完全贝叶斯的后验变异。

虽然经验贝叶斯因使用数据获得参数的先验而被批评，但是只要样本量足够大重复应用数据进行分析并不会导致问题，而这一点对 Meta 分析是满足的。相对于完全贝叶斯，经验贝叶斯往往更具有稳健性和更简洁的模型形式，而且表现出十分优良的频率学统计性质；此外，方差参数 ν^2 的先验设定也是分层模型的一个困难之处，一些常见的看似合理的无信息先验实际上却会对后验产生实质性的影响，经验贝叶斯有效地避免了类似的麻烦。

完全贝叶斯模型需要注意的事项包括：需要对指定的数值进行敏感性分析，以判断模型的稳健性；分层模型中方差参数的先验指定存在争议；后验分布没有显性表达式，需要借助采用马尔科夫蒙特卡洛算法 (Markov Chain Monte Carlo, MCMC) 进行模拟抽样，这又涉及收敛诊断的问题。采用网格法从后验分布中随机抽样，相对于其他 MCMC 算法如 Gibbs 抽样和 Metropolis-Hastings 法，网格抽样更简单，操作性强，并且不存在收敛诊断的问题。估计潜在效应 θ_i 时，由于贝叶斯分层模型认为各研究来自共同总体，因此潜在效应 θ_i 能从其他研究获得信息，估计更加精确，这称为贝叶斯分析的借力效果 (Borrow Strength Effect)。图 5-7 显示不管是经验贝叶斯还是完全贝叶斯，所有潜在效应 θ_i 的置信区间都变小了，那些小样本研究收益更多。

在实际 Meta 分析中, 纳入的研究之间或多或少存在差别, 这种差别在病例对照研究的 Meta 分析中尤为突出, 而临床试验 Meta 分析中各研究的同质性相对高一些。这是因为病例对照研究很少能做到像临床试验那样的质量控制和保证, 因此病例对照研究间显示更多的变异, 这时更有理由应用随机效应模型以考虑研究间异质性。文中的 HMD 数据来自临床试验, 从后验结果可见各研究间的同质性较好, 对 HMD 数据采用固定效应模型分析(给 ν^2 指定一个极小的数如 0.001), 得到真实效应 μ 的后验均值为 -0.467 (OR = 0.627), 对应的方差为 0.015, 总体效应和贝叶斯分层模型接近, 但因为忽略了研究间的变异, 方差变小了。

事实上, 贝叶斯统计的一个优点就在于能够直接计算感兴趣的统计量, 如可以根据 ν^2 的后验分布直接判断 ν^2 是否与 0 足够地接近, 即通过计算 ν^2 在与 0 附近的一个较小的区间(如 0 ~ 0.1)来做到这一点。从图 5-5 可见, ν^2 的后验密度大部分在 0.1 以上, $p(\nu^2 < 0.1) = 0.283$ 。这意味着, 14 个 HMD 研究试验间确实存在不可忽略的差别, 采用分层模型以考虑研究间的异质性更恰当。

另一个问题是, 假设潜在效应 θ_i 来自正态总体, 如果贝叶斯分析的诊断结果发现实际的数据违背了这一假设, 那么需要选择其他合适的总体分布。比如 t 分布是正态分布的一个稳健(Robust)替代, 能够更好地分析出现的异常研究。这种替换在频率统计下很难处理, 但贝叶斯实现却十分简单, 只需要将总体分布替换为 t 分布, 但此时后验计算变得复杂。

5.3.3 SAS 程序

1. 计数资料分析的 SAS 程序

```
data count;
input treat age use1 use2 sex count;
cards;
.....
.....
;
proc genmod data=count;
model y=treat age use1 use2 sex/dis=poisson link=log;
run;

proc genmod data=count;
model y=treat age use1 use2 sex/dis=negbin link=log;
run;
```

2. Meta 分析的 R 程序

```
options(digits=4)
library(plotrix)
set.seed(12345)
bameta=read.table("D: \\bameta.txt", sep="\t", header=TRUE)
attach(bameta)
y=log(a* d/(b* c))
s2=1/a+1/b+1/c+1/d

v2=seq(0.001, 5, 0.0001)
n=length(v2)
llv2=rep(0, n)
for (i in 1:n)
{
v=1/sum(1/(s2+v2[i]))
```

```

mu0 = v * sum(y / (s2 + v2[i]))
llv2[i] = 0.5 * sum(-log(s2 + v2[i]) - (y - mu0)**2 / (s2 + v2[i])) + 0.5 * log(v)
}

p = exp(llv2) / sum(exp(llv2))
llv = max(llv2)
mlv2 = max((llv2 == max(llv2)) * rep(1, n) * v2)

par(mfrow = c(2, 1))
plot(v2, p, type = "l", ylab = "log-likelihood", xlab = "v2", mgp = c(1, 0, 0), tcl = -0.1)
plot(v2, llv2, type = "l", ylab = "log-likelihood", xlab = "v2", mgp = c(1, 0, 0), tcl = -0.1)
list(llv, mlv2)

s = 10000
sv2 = sample(v2, s, replace = TRUE, prob = p)

beta = matrix(0, s, 14)
smu = rep(0, s)
for (i in 1:s)
#full bayesian
{
v = 1 / sum(1 / (s2 + sv2[i]))
mu0 = v * sum(y / (s2 + sv2[i]))
smu[i] = rnorm(1, mu0, sqrt(v))
mubeta = (y / s2 + smu[i] / sv2[i]) / (1 / s2 + 1 / sv2[i])
tau = 1 / (1 / s2 + 1 / sv2[i])
for (j in 1:14)
{
beta[i, j] = rnorm(1, mubeta[j], sqrt(tau)[j])
}
}

bmean = rep(0, 14)
bsb = rep(0, 14)
for (j in 1:14)
{
bmean[j] = mean(beta[, j])
bsb[j] = sd(beta[, j])
}
cbind(bmean, bsb)

#empirical bayesian
ebv = 1 / sum(1 / (s2 + mlv2))
ebmu0 = ebv * sum(y / (s2 + mlv2))
ebsmu = rnorm(s, ebmu0, sqrt(ebv))
ebbeta = matrix(0, s, 14)
for (i in 1:s)
{
for (j in 1:14)
{
ebmu = (y / s2 + ebsmu[i] / mlv2) / (1 / s2 + 1 / mlv2)
ebtau = 1 / (1 / s2 + 1 / mlv2)

```

```

ebbeta[i, j] = rnorm(1, ebmu[j], sqrt(ebtau)[j])
}
}

ebmean = rep(0, 14)
ebsb = rep(0, 14)
for (j in 1:14)
{
ebmean[j] = mean(ebbeta[, j])
ebsb[j] = sd(ebbeta[, j])
}
cbind(ebmean, ebsb)
cbind(ebmean, ebsb* ebsb, bmean, bsb* bsb)
cbind(EBL = (ebmean-1.96* ebsb), EBU = (ebmean+1.96* ebsb), FBL = (bmean-1.96*
bsb), FBU = (bmean+1.96* bsb))

par(mfrow=c(2, 2))
plot(v2[1:1500], llv2[1:1500], type="l", ylab="log-likelihood", xlab="v2",
mgp=c(1, 0, 0), tcl=-0.1)
plot(density(sv2, bw=0.05), mgp=c(1, 0, 0), tcl=-0.1, main="", xlab="v2")
plot(density(smu, bw=0.05), mgp=c(1, 0, 0), tcl=-0.1, main="", xlab="mu")
plotCI(1:14, ebmean, uiw=1.96* ebsb, liw=1.96* ebsb, lwd=2, col="blue",
scol="blue", ylim=c(-5, 3), mgp=c(1, 0, 0), tcl=-0.1, ylab="Confidence In-
terval", xlab="the number of studies")
#legend(1, 2, c("EB", "FB", "Crude"), lwd=c(2, 3, 0.5), col=c("red", "blue", "
black"), text.width=1)
plotCI(0.8:13.8, y, uiw=1.96* sqrt(s2), liw=1.96* sqrt(s2), lwd=0.5, col="
black", scol="black", add=TRUE)
plotCI(1.2:14.2, bmean, uiw=1.96* bsb, liw=1.96* bsb, lwd=3, col="red",
scol="red", add=TRUE)
cbind(list(mean(ebsmu), sd(ebsmu)** 2), list(mean(smu), sd(smu)** 2))
cbind(ebmean, bmean)

```

3. Meta 分析的 SAS 程序

```

data a;
input
subabcd;
y=log(a* d/(b* c));
s2=1/a+1/b+1/c+1/d;
cards;
1    35    498    60    476
2    1     60     5     49
3    16    117    23    117
4    3     64     7     53
5    9     48     10    61
6    3     61     13    46
7    1     70     7     68
8    4     77     11    52
9    32    340    33    338
10   5     44     2     27
11   7     114    13    111
12   1     23     1     21
13   5     31     6     31
14   6     90     8     84
;

```

```

proc mcmc data=a outpost=classout nmc=500000 thin=50 seed=1234567;
array mu[14];
parms mu: 0;
parms v 1;
prior mu: ~ normal(mean = mu0, var = v);
prior mu0 ~ normal(mean = 0, var = 1e6);
prior v ~ igamma(shape = 3/10, scale = 10/3);
model y ~ normal(mu[sub], var = s2);
run;

```

上面的程序至少需要 SAS 9.2 的环境。

参 考 文 献

- [1] 金伯泉. 医学免疫学(第5版). 北京: 人民卫生出版社, 2008.
- [2] Agresti A. An introduction to categorical data analysis, Second Edition. New York: John Wiley & Sons, 2007.
- [3] Albert J. Bayesian Computation with R, Second Edition. New York: Springer, 2009.
- [4] Alexander JS, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine*, 2008, 27: 625-650.
- [5] Berry DA, Stangl DK. Meta-Analysis in Medicine and Health Policy. New York: Chapman & Hall, 2000.
- [6] Böning D. A note on test for Poisson overdispersion. *Biometrika*, 1994, 81: 418-419.
- [7] Borenstein M, Hedges LV, Higgins JPT, *et al.* Introduction to Meta-Analysis. Chichester: John Wiley & Sons, 2009.
- [8] Brooks S, Gelman A, Jones G, *et al.* Handbook of Markov Chain Monte Carlo. London: Chapman & Hall, 2011.
- [9] Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 2006, 1: 473-514.
- [10] Cameron A, Trivedi P. Regression Analysis of Count Data. Oxford University Press, 1998.
- [11] Cameron AC, Pravin KT. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1986, 1: 29-53.
- [12] Cameron AC, Pravin KT. Regression-based tests for overdispersion in Poisson models. *Journal of the American Statistical Association*, 1990, 46: 347-264.
- [13] Cameron AC, Trivedi P. Regression Analysis of Count Data. Oxford University Press, 1998.
- [14] Carlin BP, Louis TA. Bayes and Empirical Bayes Methods for Data, Second Edition. London: Chapman & Hall, 2000.
- [15] Carlin JB. Meta-analysis for 2 x 2 tables: a Bayesian approach. *Statistics in Medicine*, 1992, 11: 141-159.
- [16] Casella G. Empirical Bayes Gibbs sampling. *Biostatistics*, 2001, 2(4): 485-500.
- [17] Congdon P. Applied Bayesian Modelling. Chichester: John Wiley & Sons, 2003.
- [18] Dean C, Lawless JF. Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, 1989, 84: 467-472.
- [19] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 1986, 7: 177-188.
- [20] Efron B. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. New York: Cambridge University Press, 2010.
- [21] Fleming TR, Harrington DP. Counting Processes and Survival Analysis. New York: Wiley, 1991.
- [22] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, Second Edition. London: Chapman & Hall, 2004.
- [23] Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 2006, 1: 515-533.
- [24] Greene W. Econometric Analysis, Fifth Edition. Prentice Hall, 2002.
- [25] Keith A, Bruno S. Approximate Bayesian inference for Random effects meta-analysis. *Statistics in Medicine*,

- 1998, 17: 201-218.
- [26] Lussenhop J. Victor Hensen and the development of sampling methods in ecology. *Journal of the History of Biology*, 1974, 7: 319-337.
- [27] Morris CN. Parametric Empirical Bayes Inference; Theory and Applications. *Journal of the American Statistical Association*, 1983, 78: 47-55.
- [28] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL <http://www.R-project.org>.
- [29] Ramon CL, George AM, Walter W, et al. SAS for Mixed Models, Second Edition. SAS Institute Inc., 2006.
- [30] Robert CP, Casella G. Monte Carlo Statistical Methods, 2nd Ed. New York: Springer-Verlag, 2004.
- [31] Spiegelhalter DJ, Thomas A, Best NG, et al. WinBUGS Version 1. 4 Users Manual. MRC Biostatistics Unit, Cambridge, 2003.
- [32] Winkelmann R. Econometric Analysis of Count Data, Fifth Edition. Berlin: Springer-Verlag, 2008.
- [33] Yusuf S, Peto R, Lewis J, Collins R et al. Beta blockade during and after myocardial infarction; an overview of the randomized trials. *Progress in Cardiovascular Diseases*, XXVII, 1985, 335-371.
- [34] 陈峰. 医用多元统计分析方法. 北京: 中国统计出版社, 2001.
- [35] 褚文政, 刘宏, 于国平等. 侧脑室注射链脲酶素导致大鼠脑内能量代谢障碍和视空间记忆障碍. *中国神经免疫学和神经病学杂志*, 2009, 6: 418-420.
- [36] 董会卿, 刘峥, 张普等. 长期应用糖皮质激素治疗对 RRMS 患者临床复发和神经功能残疾进展的影响. *中国神经免疫学和神经病学杂志*, 2010, 5: 323-326.
- [37] 方积乾, 陆盈. 现代医学统计学. 北京: 人民卫生出版社, 2002, 150-206.
- [38] 胡昱, 陈雅萍, 李倩等. Oka 株水痘减毒活疫苗流行病学保护效果 Meta 分析. *中国疫苗和免疫*, 2011, 5: 431-435.
- [39] 胡昱, 李倩, 陈恩富等. 口服轮状病毒减毒活疫苗保护效果研究 Meta 分析. *中国疫苗和免疫*, 2012, 1: 61-66.
- [40] 胡昱, 李倩, 蒋征刚等. 国内流行性感胃灭活疫苗的保护效果研究 Meta 分析. *中国疫苗和免疫*, 2011, 4: 343-349.
- [41] 罗顺元, 杨婧, 陈梅桂. 针灸治疗类风湿关节炎的 Meta 分析. *中国卫生统计*, 2009, 26(4): 431-433.
- [42] 孙庆利, 樊东升, 孙阿萍等. 中国多发性硬化患者肿瘤坏死因子-基因多态性的 Meta 分析. *中国神经免疫学和神经病学杂志*, 2011, 4: 250-255.
- [43] 覃莲香, 吕继成. 干扰素调节因子 5 基因 rs2004640 多态性与系统性红斑狼疮遗传易感性的荟萃分析. *中国免疫学杂志*, 2011, 27(3): 239-242.
- [44] 肖华, 陈志坚, 廖玉华等. TNF- α 早期表达在大鼠急性心肌梗死室性心律失常发生中的作用及机制研究. *中国免疫学杂志*, 2008, 24(1): 75-78.
- [45] 杨娟, 郑青山. Meta 分析的统计学方法. *中国临床药理学与治疗学*. 2005, 10(11): 1309-1314.
- [46] 张丽, 许继超, 荣爽等. 莲房原花青素对东莨菪碱所致小鼠记忆获得性障碍的改善作用. *中国神经免疫学和神经病学杂志*, 2009, 6: 406-410.
- [47] 张敏华, 吴金锴, 应勇. 西洛司特对 COPD 患者炎症水平的调节作用. *放射免疫学杂志*, 2011, 1: 67-68.
- [48] 张枢, 王宇, 张宇. 艾叶挥发油治疗大鼠变应性鼻炎的试验研究. *中国免疫学杂志*, 2011, 9: 787-789.
- [49] 赵风源, 贺圣文, 赵仁宏等. 胃癌危险因素 Meta 分析. *中国卫生统计*, 2010, 27(2): 146-148.
- [50] 赵秀娟, 张卫红, 孙素娟等. ω -3 多不饱和脂肪酸对老年痴呆大鼠学习记忆和海马 MDA、SOD 的影响. *中国神经免疫学和神经病学杂志*, 2009, 16(4): 273-275.
- [51] 丁元林. 卫生统计学. 北京: 科学出版社, 2010.
- [52] 陆守曾, 陈峰. 医学统计学(第2版). 北京: 中国统计出版社, 2007.

第 6 章 高血压病统计设计与数据分析

《中华高血压杂志》是国内唯一的高血压及相关疾病诊疗、防治和科研的医学专业期刊，主要交流国内外高血压及相关疾病的临床防治经验与科研成果。设有编辑部述评、学术争鸣、指南、受体阻滞剂专栏、综述、论著(临床、基础、预防医学)、临床研究快讯、临床经验交流、讲座、病例分析、中美互换摘要、国内外动态等栏目。

本章内容包括该杂志统计方法应用、统计设计、统计分析中常见问题及案例辨析。

6.1 论著统计方法应用情况

《中华高血压杂志》2008 年和 2011 年发表论著共 280 篇，未使用统计分析方法的仅有 3 篇，统计分析方法的使用率达 98.9%。280 篇论著中试验研究有 162 篇，观察性研究有 118 篇，各种统计方法的使用情况见表 6-1。由表 6-1 可知，在 280 篇论著中 t 检验，方差分析和 χ^2 检验的使用频率比较高，是常用的统计分析方法。

表 6-1 280 篇论著中统计方法的应用情况(篇)

年 份	t 检 验	方 差 分 析	χ^2 检 验	非参数检验	相 关	回 归	其 他	未 使 用
2008 年	66	79	36	3	21	19	3	2
2011 年	51	56	55	12	29	41	12	1

注：其他包括协方差分析，一致性检验，荟萃分析，因子分析等。

280 篇论著中 2008 年有统计学错误的 59 篇，2011 年有 47 篇。其错误类型见表 6-2。2008 年主要为统计分析(54.2%)方面的错误，2011 年主要为统计结果表达与描述(51.1%)的错误。

表 6-2 280 篇论著中存在的统计学错误(篇)

年 份	n	试 验 设 计	统 计 分 析	统 计 表 达
2008 年	59	17	32	14
2011 年	47	8	18	24

280 篇论著在试验设计方面的错误分布情况见表 6-3。试验设计在整个研究过程中起着举足轻重的作用，如果试验设计出现问题将不能保证试验结果的科学性。

表 6-3 208 篇论著试验设计的统计学错误(篇)

错 误 类 型	2008 年	2011 年	错 误 类 型	2008 年	2011 年
无对照或对照不合理	6	1	受试对象不同质或同质性差	2	0
违背均衡性原则	1	1	样本含量太少	5	3
违背随机原则	2	2	违背重复原则	1	1

6.2 试验设计中的常见问题及案例辨析

调查了 2008 年和 2011 年《中华高血压杂志》的统计应用情况，结果表明不合理应用率为 37.86%。试验设计是整个试验过程中的重要环节，良好的研究设计是顺利进行科学研究和数据分

析的先决条件,也是获得预期结果的重要保证。统计学的使用应该贯穿于整个科研工作的始终,而不是仅仅局限于最后的试验数据处理(分析)阶段。

对照是试验设计的重要原则之一,合理的对照可以排除或减少非处理因素的干扰。在试验设计过程中,其常见的问题有:①缺少对照组;②对照组不恰当。如对照组的例数太少,组间缺乏可比性。

随机化原则是避免偏倚和控制混杂因素最有效的方法。常见的问题有:①未指出分组是否随机,或是否采用随机化的分组方法;②把随意当作随机,按入院顺序、单双日等分组。

此外,样本的选择在试验研究中也是非常重要的。有时研究者在试验设计中选取的样本缺乏代表性,不能够很好地代表研究总体,影响结果的外推。此外,样本含量也是一个重要问题,样本只有足够大且对总体具有极好的代表性才可以对所提出的问题作出可靠的回答。

【案例6-1】 Tol 样受体4(TLR4)的激活与动脉粥样斑块的进展及斑块的不稳定导致临床并发症的发生有关。他汀类调脂药物临床获益可能与其调脂以外的作用尤其是抗炎作用有关,但抗炎作用的确切机制目前还不清楚。为了推测其抗炎作用可能是通过抑制 TLR4 炎症信号通路起作用的,选取了健康志愿者($n=22$)及2006年7月至2007年9月某医院门诊和住院患者共121例(稳定性心绞痛(SAP)患者17例,急性冠脉综合征(ACS)患者82例),入院即刻抽取外周静脉血。将其中的41例ACS患者随机分为两组:在常规抗心绞痛治疗的基础上服用阿托伐他汀10mg组(20例)及阿托伐他汀40mg组(21例),治疗1月后抽取外周静脉血。观察指标为血脂、高敏C反应蛋白(hsCRP)、外周血CD4⁺及单核细胞表面TLR4的表达,单核细胞表面TLR4的表达采用流式细胞仪方法测定。文章部分数据见表6-4。

表6-4 两组患者用不同剂量阿托伐他汀治疗前后血脂的变化($\bar{x} \pm s$)

阿托伐他汀 剂量组	例数	TC (mmol/L)			TG (mmol/L)		
		治 疗 前	治 疗 后	变化(\pm)	治 疗 前	治 疗 后	变化(\pm)
10mg	20	4.56 \pm 1.14	3.57 \pm 0.98 ^a	(-)0.81 \pm 0.45	2.53 \pm 1.02	1.89 \pm 0.61	(-)0.64 \pm 0.75
40mg	21	4.85 \pm 1.17	3.27 \pm 0.59 ^b	(-)1.58 \pm 0.87 ^c	1.91 \pm 1.28	1.18 \pm 0.49	(-)0.72 \pm 1.06
阿托伐他汀 剂量组	例数	LDL-C (mmol/L)			HDL-C (mmol/L)		
		治 疗 前	治 疗 后	变化(\pm)	治 疗 前	治 疗 后	变化(\pm)
10mg		2.71 \pm 0.99	2.22 \pm 0.91 ^a	(-)0.49 \pm 0.29	1.01 \pm 0.09	1.02 \pm 0.11	(+)0.01 \pm 0.16
40mg		3.08 \pm 0.94	1.99 \pm 0.48 ^a	(-)1.09 \pm 0.65 ^b	1.01 \pm 0.20	1.07 \pm 0.21	(+)0.06 \pm 0.21

注: TC: 总胆固醇; TG: 三酰甘油; LDL-C: 低密度脂蛋白胆固醇; HDL-C: 高密度脂蛋白胆固醇; 与治疗前比较, a: $P < 0.05$, b: $P < 0.01$; 与40mg组治疗前后的变化比较, c: $P < 0.05$ 。

【案例辨析】 原作者把服用阿托伐他汀前急性冠脉综合征(ACS)患者的基线资料作为自身对照,用以说明阿托伐他汀的药效。但是自身对照不能够完全说明药物的药效,因为在试验过程中给予了受试对象药物,受试者由于心理作用会使药物的作用被放大。

【案例释疑】 如果该研究在伦理道德和患者知情同意方面没有问题的话,可考虑增设一个安慰剂对照组。有比较才有鉴别,对照是比较的基础。合理设置对照组可以排除或减少非处理因素的干扰和影响,使不同处理因素所产生的效应有科学的对比。

【案例6-2】 为了观察阿托伐他汀对无脂代谢紊乱的高血压患者外周血内皮祖细胞数量及血压的影响。选择原发性高血压患者38例随机分为单用常规降压药物组(常规组, $n=18$)和常规降压药物与阿托伐他汀(20mg 睡前)联合用药组(联合组, $n=20$),8例健康志愿者口服阿托伐他汀作为对照组(对照组, $n=8$)。于治疗前和治疗后8周分别测血压并抽取外周血进行EPCS的分离培养,第10天对EPCS进行鉴定并于倒置相差显微镜下计数内皮祖细胞克隆形成单位(EPC-CFU)以评估外周血EPCS水平。文章部分数据见表6-5。

表 6-5 3 组临床资料的比较 ($\bar{x} \pm s$)

组别	例数 (男/女)	年龄 (岁)	吸烟 (人)	饮酒 (人)	应用降压药物的例数			
					β 受体阻滞剂	钙拮抗剂	ACEI 或 ARB	利尿剂
常规	18(9/9)	62.5 \pm 8.7	7	6	7	15	8	12
联合	20(11/9)	61.1 \pm 8.2	8	8	8	16	9	14
对照	8(4/4)	63.4 \pm 7.8	3	3	0	0	0	0

注：ACEI，血管紧张素转换酶抑制剂；ARB，血管紧张素Ⅱ受体拮抗剂。

【案例辨析】 原作者将健康志愿者作为对照组，忽视了原发性高血压患者同健康人群在药物代谢和吸收方面的差别。正常健康人和原发性高血压患者在各项生理功能上是有差别的，而且药物在联合使用时有可能会产生各种不同的效应，如协同，拮抗等。

【案例释疑】 正确的做法是建立原发性高血压患者单独使用阿托伐他汀的对照组，并与常规药物进行比较。如果差异有统计学意义则可认为阿托伐他汀对无脂代谢紊乱的高血压患者外周血内皮祖细胞数量及血压的影响。

【案例 6-3】 为了研究社会地位和教育水平较高人群的血压和血脂水平，作者选用一组高校教师作为研究对象，通过整群抽样的方法选取 1028 名高校教师，当地的 862 例市直属企业单位健康体检人群作为对照，测量其血压、身高、体质质量指数、血脂并进行比较，得出高校教师的血压和血脂水平同普通人的血压和血脂水平差异有统计学意义。文章部分数据见表 6-6、图 6-1、图 6-2。

表 6-6 高校教师人群和普通人群身高、体质质量比较

调查对象	例数	年龄(岁)	身高(cm)	体质质量(kg)	体质质量指数(kg/m ²)
教师人群	1028	42.3 \pm 14.4	154.1 \pm 7.9	51.1 \pm 7.3	21.4 \pm 2.7
普通人群	862	46.8 \pm 15.1	153.8 \pm 8.0	52.3 \pm 7.8 ^a	22.3 \pm 2.5 ^a

注：与高校教师人群比较，a：P < 0.01。

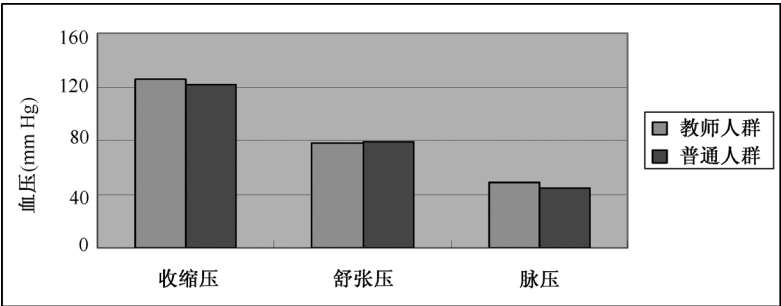


图 6-1 高校教师人群和普通人群血压水平比较

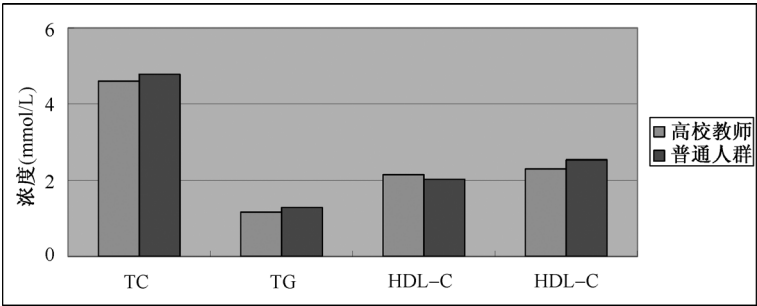


图 6-2 高校教师人群和普通人群血脂水平比较

【案例辨析】 社会地位与血压和血脂水平的关系在研究过程中未提及,且高校教师只是代表了教育水平较高者的一部分。并且在对照时,原作者将市直企事业单位健康体检人群选为对照当作普通人,其代表性不强。

【案例释疑】 选取社会地位和教育水平较高的多种职业人群作为研究对象,对照组在选择时可采用分层抽样选取多种职业人群。

【案例6-4】 饮食治疗和生活方式的改善始终是临床预防和干预的基础。为了探讨不同膳食结构对原发性高血压患者心血管病危险因素的影响,应用膳食调查问卷形式研究者连续选取共424例2005年1月至2006年12月某院高血压门诊的患者,均为初诊的原发性高血压患者,或者虽已诊断原发性高血压但未规律服用降压药物及调脂药物,其中男性261例,女性163例,年龄(48.8 ± 10.2)岁。测量患者的身高、体质量、腰围,并清晨空腹取静脉血检测各项生化指标。然后将入选对象分为三组:素食组、肉食组、平衡组。文章部分数据见表6-7。

表6-7 3组原发性高血压患者之间临床及生化指标的比较

组 别	例 数	年龄 (岁)	BMI	腰围 (cm)	高血压史 (年)	收缩压 (mmHg)	舒张压 (mmHg)
素食	95	52.0 ± 10.6	25.1 ± 3.2	87.4 ± 0.9	6.4 ± 8.6	153.3 ± 19.5	98.2 ± 11.0
肉食	133	45.8 ± 9.5 ^a	27.9 ± 3.4 ^a	96.5 ± 9.4 ^a	5.8 ± 7.4	147.0 ± 19.8 ^b	98.1 ± 14.7
平衡	196	49.5 ± 10.5	26.3 ± 3.4 ^{ac}	91.4 ± 10.1 ^{ac}	7.2 ± 9.9	149.8 ± 20.7	95.2 ± 13.4

组 别	脉压 (mmHg)	心率 (次/min)	ALT (IU/L)	AST (IU/L)	肌酐 (mmol/L)	空腹血糖 (mmol/L)	UA (mmol/L)
素食	55.2 ± 15.1	72.1 ± 9.4	25.4 ± 18.5	26.2 ± 12.1	75.9 ± 17.9	5.6 ± 1.9	268.3 ± 79.3
肉食	49.0 ± 14.7 ^a	74.9 ± 12.8	35.9 ± 23.2 ^a	28.9 ± 11.5	91.1 ± 19.6 ^a	5.7 ± 1.5	335.4 ± 85.4 ^b
平衡	54.7 ± 17.1 ^a	73.8 ± 11.8	25.1 ± 17.6 ^c	24.8 ± 8.4 ^c	84.8 ± 18.8 ^{ac}	5.6 ± 1.9	292.7 ± 85.4 ^{ac}

组 别	TG (mmol/L)	TC (mmol/L)	HDL-C (mmol/L)	LDL-C (mmol/L)	血清钾 (mmol/L)	血清钠 (mmol/L)
素食	2.2 ± 1.8	4.9 ± 0.8	1.3 ± 0.3	2.9 ± 0.6	4.1 ± 0.5	140.7 ± 2.7
肉食	3.2 ± 2.7 ^a	5.0 ± 1.1	1.1 ± 0.2 ^a	2.9 ± 0.9	4.0 ± 0.4	141.0 ± 2.9
平衡	2.3 ± 2.2 ^c	5.0 ± 1.1	1.2 ± 0.3 ^c	3.0 ± 0.8	4.1 ± 0.4	140.6 ± 3.1

注: BMI: 体质量指数; ALT: 丙氨酸氨基转移酶; AST: 天冬氨酸氨基转移酶; TG: 三酰甘油; TC: 总胆固醇; HDL-C: 高密度脂蛋白胆固醇; LDL-C: 低密度脂蛋白胆固醇; 与素食组比较, a: $P < 0.05$, b: $P < 0.01$; 与肉食组比较, c: $P < 0.01$ 。

【案例辨析】 在选取研究对象时采用的是连续选取,这样看起来是随机的,其实,分入各组的患者可能在年龄、性别等某些重要的非处理因素上相差很多。从表6-7可以看出素食组在年龄上与肉食组和平衡组的差别有统计学意义。也就是说各组之间的差别很可能是在研究对象的选取过程中引起的。没有通过“随机化”方法尽量降低重要非试验因素对观测结果的影响,同时也就严重地违背了试验设计中的“均衡原则”,其试验结果的可靠性必然要受到影响。

【案例释疑】 在研究对象选取时可按性别和年龄进行分层抽样。

【案例6-5】 采用动态血压监测(ABPM)观察比索洛尔/氢氯噻嗪复方片对轻、中度原发性高血压患者服药后18~24h血压和血压晨峰(MBPS)程度的影响。选择90例轻、中度高血压患者[收缩压(SBP) $< 180\text{mmHg}$, 舒张压(DBP) $95 \sim 109\text{mmHg}$], 口服比索洛尔/氢氯噻嗪复方片2.5mg/6.25mg, 4周后血压不达标者用比索洛尔/氢氯噻嗪复方片5.0mg/6.25mg继续治疗4周,应用ABPM评价治疗前后服血压和血压晨峰程度的变化。

【案例辨析】 研究者在研究不同浓度比索洛尔/氢氯噻嗪复方片对血压晨峰的影响时,高浓度组(5.0mg/6.25mg)的研究对象是低浓度组中(2.5mg/6.25mg)治疗4周后血压不达标者。这种研究在试验设计上是不合理的,高浓度组的研究对象在试验过程中同时接受两种不同浓度的药物治

疗,经过低浓度组的筛选后剩下的都是比较耐药的。由于缺乏相应的对照,高浓度组对血压晨峰的影响得不到很好的反映。

【案例释疑】 在设计时应该同等考虑各种浓度的效应,单独设置一个高浓度组。

6.3 统计分析中的常见问题及案例辨析

6.3.1 偏态分布误按正态分布处理

【案例6-6】 目的:分析原发性醛固酮增多症(PA)的临床表现、辅助检查、治疗及术后血压变化。数据处理应用SPSS 16.0统计软件分析完成,定量资料以均数 \pm 标准差($\bar{x} \pm s$)表示,采用 t 检验,以 $P < 0.05$ 为差异有统计学意义。得出APA和IHA两组的性别、病程、血压、血钾、24 h尿钾、立卧位醛固酮、立卧位肾素活性、立卧位ARR无统计学意义($P > 0.05$)。

【案例辨析与释疑】 文章说明定量资料以均数 \pm 标准差($\bar{x} \pm s$)表示,并没有考虑资料的实际分布。正态分布或者大样本近似正态分布资料用均数 \pm 标准差($\bar{x} \pm s$)表示,采用 t 检验;偏态分布资料用中位数(四分位间距)即 $M(Q_R)$ 表示,用秩和检验。病程为偏态资料,而文中APA和IHA样本含量分别是33和75,样本含量较少并不能近似正态处理,用秩和检验来比较两组病程差异更为合适。

6.3.2 未标明所采用的统计分析方法

【案例6-7】 目的:探讨我国中年人群血压状态的自然转归和进展,为确定高血压的防治策略提供依据。表6-8显示不同服药状态下随访和失访人群的基本特点。与失访人群比较,随访人群除年龄、BMI和女性所占比例较高外,血压水平(SBP/DBP)、体质量、饮酒率、体育锻炼比例和高血压的患病率均没有显著差异($P > 0.05$)。

表 6-8 研究对象两次调查时基本特征($\bar{x} \pm s$)

项 目	基 线 项 目								基线血压分类(%)				
	年龄 (岁)	SBP (mmHg)	DBP (mmHg)	体质量 (kg)	BMI (kg/m ²)	饮酒 (%)	吸烟 (%)	体育锻 炼(%)	正常	正常 高值	轻度 高血压	中重度 高血压	
基线服药	51.0±5.8	140.9±18.5	87.8±9.3	66.8±12.5	26.1±3.7	17.6	25.0	56.9	—	—	41.2	58.8	
复查服药	49.6±6.7	144.6±22.2	89.6±12.4	63.6±11.0	24.7±3.5	21.9	30.4	39.8	10.1	24.6	33.7	31.5	
两次均服药	51.1±6.3	152±21.2	93.6±11.3	67.9±10.5	26.2±3.6	12.3	32.3	56.8	—	—	20.9	79.1	
两次均未服药	46.5±7.0	119.0±15.7	75.6±10.2	58.5±10.3	22.8±3.2	25.5	33.9	28.4	49.8	36.2	11.1	2.9	
随访	47.0±7.0	122.9±19.3	77.7±11.7	59.4±10.7	23.1±3.4	24.6	33.4	30.7	43.7	33.4	13.9	8.9	
未复查	46.6±7.5	122.3±19.5	77.8±12.0	59.8±10.4	22.9±3.3	24.3	39.7	37.5	43.8	34.6	13.7	8.0	
项 目	人数	女性 (%)	复 查 项 目						复查血压分类(%)				
			SBP (mmHg)	DBP (mmHg)	体质量 (kg)	BMI (kg/m ²)	饮酒 (%)	吸烟 (%)	体育锻 炼(%)	正常	正常 高值	轻度 高血压	中重度 高血压
基线服药	51	64.7	144.5±25.2	85.8±13.0	65.0±13.7	25.5±4.3	17.6	25.5	55.8	17.6	21.6	35.3	25.5
复查服药	593	59.0	145.9±20.0	87.8±11.6	65.4±11.2	25.5±3.6	24.6	29.5	39.6	—	—	33.4	66.6
两次均服药	220	59.1	149.4±23.4	88.0±11.1	67.7±11.0	26.2±3.5	18.6	30.5	57.2	—	—	32.3	67.7
两次均未服药	5265	54.0	123.6±17.7	78.5±10.6	59.6±10.9	23.2±3.4	28.0	34.4	29.5	38.2	39.8	16.3	5.7
随访	6129	54.8	126.9±20.0	79.8±11.3	60.5±11.1	23.6±3.5	27.3	33.7	31.8	33.0	34.3	18.7	14.0
未复查	4310	48.6	—	—	—	—	—	—	—	—	—	—	—

【案例辨析与释疑】文中未标明使用了哪种统计分析方法,而直接得出 p 值,作出结论,缺乏说服力。根据原文,要进行两组人群基本特征的比较,应首先考察各项的资料类型及分布特征,选择合适的检验方法。如年龄、BMI、血压水平(SBP/DBP)、体质量为定量资料,若满足 t 检验条件(独立性、正态性、方差齐性),求得统计量 t 值,确定 p 值,与检验水准比较得出结论。如果不满足 t 检验条件,则采用校正 t 检验或进行变量变换或非参数检验。而女性所占比例、饮酒率、体育锻炼比例和高血压的患病率可以用 χ^2 检验或Fisher精确检验来分析。

6.3.3 误用统计描述结果作出统计推断解释

【案例6-8】目的:探讨我国中年人群血压状态的自然转归和进展,为确定高血压的防治策略提供依据。表6-8显示不同服药状态下随访和失访人群的基本特点。结果:与基线相比,除两次调查均服药者外,其他研究对象6年后血压均升高(SBP/DBP)4/2 mmHg。

【案例辨析与释疑】表6-8为统计描述,直接比较所得结论显然不可靠,要知道6年后血压是否升高,应该进行假设检验。并且表中资料并不能看出与基线相比,除两次调查均服药者外,其他研究对象6年后血压均升高(SBP/DBP)4/2 mmHg。

6.3.4 误用单因素方差分析处理重复测量资料

【案例6-9】为观察苯那普利、厄贝沙坦对慢性心力衰竭(CHF)大鼠主动脉内皮氧化应激以及一氧化氮合酶(NOS)、还原型辅酶(NADPH)、氧化酶重要亚单位p22phox蛋白表达的影响。采用大鼠腹主动脉缩窄法造成压力负荷性心肌肥厚致心力衰竭模型,分为5组:假手术组、未治疗组、苯那普利组、厄贝沙坦组及联合用药组。每2周测定大鼠尾动脉血压。采用单因素方差分析,得出结论:整个试验过程中各治疗组之间血压无差异。

【案例辨析与释疑】重复测量是对同一研究对象的同一测量指标或多个测量指标,在不同时间点(或同一个体不同部位等)的多次测量结果。该研究每两周测量大鼠动脉血压,研究各组之间血压变化是否有差异,比较不同点大鼠尾动脉血压的变化趋势,为重复测量设计,不满足单因素方差分析样本独立性条件,应采用重复测量设计方差分析。

6.3.5 误用随机区组设计方差分析处理完全随机设计资料,并进行两两比较

【案例6-10】为研究坎地沙坦对高脂饮食所致肥胖大鼠胰岛素敏感性的影响及其可能机制。将4~6周龄雄性Wistar大鼠随机分为普通饲料组(NC, $n=15$)、高脂饲料组(HF, $n=15$)和高脂饲料加坎地沙坦药物组(HF+C, $n=15$)。药物组每天灌胃坎地沙坦西酯8 mg/kg,其他两组分别给予等量生理盐水。喂养16周后收集数据,采用配伍组设计资料的方差分析,以 $P<0.05$ 为差异有统计学意义。各组大鼠体质量及脏器质量的变化(表6-9):HF、HF+C组大鼠体质量、附睾、肾周脂肪、肝脏质量显著高于NC组($P<0.01$),HF+C组心脏质量显著低于NC组($P<0.01$),而HF组与NC组差异无统计学意义。HF+C组和HF组相比,各组差异均有统计学意义。

表6-9 试验大鼠16周龄时体质量及部分脏器质量($\bar{x} \pm s$)

组别	体质量(g)	肾周脂肪质量(g)	辜丸脂肪质量(g)	肝脏质量(g)	心脏质量(g)
NC	461.3 ± 36.5	13.5 ± 3.2	7.6 ± 1.7	13.0 ± 1.6	1.34 ± 0.07
HF	518.4 ± 28.3	26.9 ± 4.8	11.5 ± 2.0	17.9 ± 1.8	1.29 ± 0.13
HF+C	478.3 ± 34.8	20.5 ± 4.1	9.6 ± 1.8	15.7 ± 1.7	1.09 ± 0.11

【案例辨析与释疑】文章将4~6周龄雄性Wistar大鼠随机分为3组,比较大鼠16周龄时三组大鼠体质量及脏器质量的变化。为完全随机设计(注意:属于非标准的单因素三水平设计,正确做法应拆分三组,第一组与第二组可比;第二组与第三组可比)的资料,近似采用完全随机设计定量资料一

元方差分析。分析结果只能说明三组差异是否有统计学意义,不能说明三组中任意两组是否有差异。不能得出结论 HF、HF + C 组大鼠体质量、附睾、肾周脂肪、肝脏质量显著高于 NC 组($P < 0.01$), HF + C 组心脏质量显著低于 NC 组($P < 0.01$),而 HF 组与 NC 组差异无统计学意义。要进行两两比较,可采用 SNK 检验(即 q 检验)、LSD 检验(即最小显著差检验)等。

严格地说,本例缺少了一组“NC + C”,属于“对照不全”的设计!若加上这样一组,就构成一个 2×2 析因设计。此时,可采用两因素析因设计定量资料一元方差分析处理资料,不仅可以分析两个试验因素(饮食类型、坎地沙坦药物用否)对各个定量指标的影响如何,还可分析两试验因素之间的交互作用是否具有统计学意义。

6.3.6 误用 t 检验和单因素多水平设计定量资料方差分析处理具有一个重复测量两因素设计定量资料

【案例 6-11】 使用链脲霉素腹腔注射法诱导雄性小鼠糖尿病,在造模成功后第 2 周将糖尿病小鼠随机分为糖尿病组及氨基胍组,其中氨基胍组小鼠用氨基胍水溶液灌胃,正常组小鼠腹腔注射等量柠檬酸-柠檬酸钠缓冲液,饲养 8 周后 3 组小鼠进行左侧股动脉高位结扎离断术造成单侧后肢缺血模型,通过红四氮唑染色法与后肢血管造影确定造模成功。利用 CD31 免疫组化染色评估股动脉结扎术后 0、1、3、7、14 及 28d 小鼠双后肢血管新生情况,观察糖尿病动物组织缺血后血管新生情况以及糖基化终产物(AGEs)形成抑制剂氨基胍对糖尿病小鼠组织缺血后血管再生障碍的影响。所有统计学分析由 SPSS 软件完成,定量资料均以均数 \pm 标准差($\bar{x} \pm s$)表示,多组均值比较采用单因素方差分析或双因素方差分析,采用 Bonferroni 检验进行均值间的两两比较,以 $P < 0.05$ 为差异有统计学意义。研究结果表明糖尿病小鼠基础毛细血管密度明显低于正常组,组织缺血后糖尿病小鼠血管新生障碍,氨基胍可部分改善糖尿病组织缺血后的血管新生。

【案例辨析】 原作者在研究氨基胍对糖尿病小鼠组织缺血后血管新生障碍的改善作用时,建立雄性小鼠糖尿病模型,采取不同的处理措施,分为正常组、糖尿病组和氨基胍组三组,并建立组织缺血模型,于缺血模型建立后的 0、1、3、7、14 及 28d 观察小鼠双后肢血管新生情况,将三组同时时间的双侧后肢腓肠肌组织的毛细血管密度进行单因素方差分析和 t 检验的两两比较,显然有误。文章部分数据见表 6-10。

表 6-10 小鼠试验前后体质量与血糖变化情况($\bar{x} \pm s$)

组 别	体质量(g)			血糖(mmol/L)		
	试 验 前	手 术 前	处 死 前	试 验 前	手 术 前	处 死 前
正常	25.0 \pm 1.1	31.5 \pm 2.5 ^a	29.6 \pm 2.0 ^b	6.7 \pm 0.9	6.7 \pm 1.1	6.8 \pm 0.5
糖尿病	25.2 \pm 1.3	26.2 \pm 1.4 ^b	23.3 \pm 1.3 ^b	6.7 \pm 1.1	20.4 \pm 2.3 ^{ab}	22.5 \pm 2.4 ^{ab}
氨基胍	25.1 \pm 1.2	27.5 \pm 1.8 ^b	25.9 \pm 1.5 ^b	6.8 \pm 0.6	18.7 \pm 1.3 ^{ab}	20.4 \pm 1.6 ^{ab}

注:与试验前比较, a: $P < 0.05$; 与正常组比较, b: $P < 0.05$ 。

【案例释疑】 该资料试验设计涉及两个因素:时间和分组因素,从原文判断试验因素时间是个典型的重复测量因素,该资料属于具有一个重复测量两因素设计定量资料,因此该资料不能采用单因素多水平设计资料方差分析,而必须采用具有一个重复测量两因素设计定量资料一元方差分析。

【案例 6-12】 将 DSS 高血压大鼠随机分为 4 组:正常饮食组,高盐饮食组(高盐组),高盐饮食辅以 NAC 治疗组(高盐 + NAC 组),正常饮食辅以 NAC 治疗组(正常饮食 + NAC 组),每组 6 只。每周测定鼠尾动脉血压,共 4 周。收集 24 h 尿,处死后取主动脉组织标本。DCF 荧光法检测尿中氧化应激生物标志物过氧化氢(H_2O_2)水平;光泽精化学发光法检测主动脉超氧阴离子生成量;RT-PCR 及 Western Blot 检测主动脉 Rho A、ROK mRNA 及蛋白表达。DSS 高血压大鼠主动脉

ROK 表达增加,表明 Rho 信号参与盐敏感性高血压的发生发展。定量数据以均数 \pm 标准差 ($\bar{x} \pm s$) 表示,采用方差分析作统计学分析(应用 StatView 软件),并作 Scheffe post hoc 多样本重复比较,以 $P < 0.05$ 为差异有统计学意义。研究结果表明 RhoA 可能不是此类高血压动物 ROK 激活的决定因素。NAC 阻断后 ROK 表达显著下调,其机制可能与 ROS 的诱导作用降低有关,ROS 可能作为上游信号直接参与 DSS 高血压大鼠 ROK 的激活。

【案例辨析】原作者为研究氧化应激对 Dahl 盐敏感性高血压大鼠主动脉 Rho 激酶的直接作用,将 DSS 高血压大鼠随机分为 4 组:正常饮食组,高盐饮食组(高盐组),高盐饮食辅以 NAC 治疗组(高盐 + NAC 组),正常饮食辅以 NAC 治疗组(正常饮食 + NAC 组)。四周后采用方差分析和 Scheffe post hoc 多样本重复比较各组动脉血压、尿 H_2O_2 、主动脉超氧阴离子、主动脉组织 Rho A 和 Rho 激酶 mRNA 表达、主动脉组织 Rho 激酶蛋白表达水平,显然有误。文章部分数据见图 6-3。

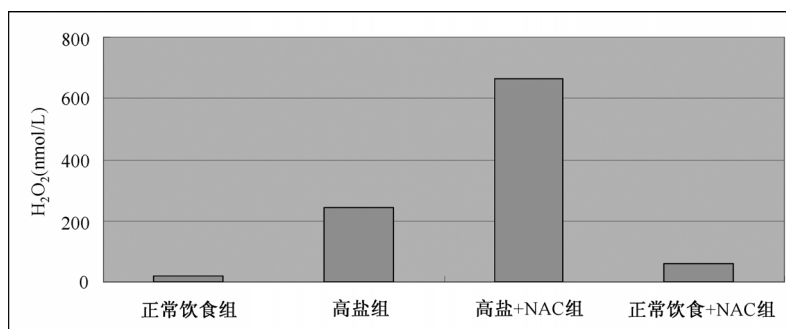


图 6-3 4 组 DSS 高血压大鼠 24h 尿中 H_2O_2 排出量

【案例释疑】该资料试验设计涉及三个因素:处理的时间、是否高脂饮食、是否使用 NAC 药物,从原文可判断试验因素处理的时间是个典型的重复测量因素,该资料属于具有一个重复测量三因素(一个时间因素、是否高脂饮食、是否使用 NAC 药物)设计定量资料,因此该资料不能采用单因素多水平设计定量资料一元方差分析和 Scheffe post hoc 多样本多重比较,而必须采用具有一个重复测量三因素设计定量资料一元方差分析来进行处理。

参考文献

- [1] 万献尧, 张久之. 医学科研或论文中常见的统计学问题. 医学与哲学, 2009, 30 (12): 74-75.
- [2] 张立群, 魏丽惠. 医学科研论文中常见统计学错误分析. 中国妇产科临床杂志, 2006, 11(7): 473-474.
- [3] 章新生, 刘璐, 姚仁斌等. 医学稿件中统计学分析使用调查. 中国科技期刊研究, 2011, 22 (3): 393-395.
- [4] 刘清海, 方积乾. 医学期刊统计学误用现状、趋势与对策. 中国科技期刊研究, 2006, 17(4): 549-551.
- [5] 毛宗福. 105 篇临床论著主要统计学缺陷调查分析. 医学新知杂志, 1997, 7(2): 92-94.
- [6] 王绍贤, 史凡, 王滨燕. 中华医学会五种主要杂志 1984 年论文中常见的统计问题分析. 中华医学杂志, 1986, 66 (9): 548-554.
- [7] 王倩, 张博恒. 五种中华医学会系列杂志论著中统计方法的应用现况. 中华医学杂志, 1998, 78(3): 230-233.
- [8] 许小幸, 蔡虹蔚, 陈光榆, 曹立明. 《临床儿科杂志》近五年论著中统计学应用情况及思考. 临床儿科杂志, 2011, 29 (1): 96-97.
- [9] 刘桂芬主编. 医学统计学 (第 2 版). 北京: 中国协和医科大学出版社, 2007.
- [10] 胡良平, 李子建. 医学统计学基础与典型错误辨析. 北京: 军事医学科学出版社. 2003.

第7章 病理研究中的统计设计与数据分析

《中华病理学杂志》主要收录了有关病理学多个方面的内容，如人体病理学、试验病理学、细胞病理学、标准与规范、病例报告等。

本章针对《中华病理学杂志》近5年发表论文中科研设计与统计分析出现的问题进行辨析与释疑，包含了研究设计、定量资料统计分析、定性资料统计分析、多因素分析及书写和概念等方面的错误，为广大病理学科研工作者提供参考。

7.1 杂志近5年论文分析

统计了《中华病理学杂志》2007年至2011年发表文章情况，五年间共刊登文章1387篇，其中人体病理学方面论著366篇；试验或细胞病理学方面论著116篇；病例报告有277篇，见表7-1。

表7-1 2007—2011年《中华病理学杂志》刊登论文篇数

论文分类	篇数(篇)				
	2007年	2008年	2009年	2010年	2011年
人体病理学	68	68	79	79	72
试验/细胞病理学	33	20	17	24	22
病例报告	53	52	55	59	58

《中华病理学杂志》近5年刊登的文章中涉及统计学内容的有375篇，约占1/4；采用某种统计推断方法的约占375篇的2/3。按照统计推断方法分类，大多数文章采用的是基本统计推断方法，如 t 检验、方差分析、 χ^2 检验、秩和检验、简单相关分析(表7-2)；有少部分文章涉及Logistic回归分析、Cox回归分析、Kaplan-Meier生存分析、Log-rank检验、Kappa一致性分析、灵敏度和特异度等分析。

表7-2 2007—2011年《中华病理学杂志》刊登论文统计方法汇总

项目		篇数(篇)					
		2007年	2008年	2009年	2010年	2011年	合计
统计学方法	有	68	69	79	87	72	375
	无	222	212	207	173	198	1012
统计推断	有	37	47	54	56	49	243
	无	31	22	25	31	23	132
t 检验/方差分析		13	10	12	15	11	61
χ^2 检验		26	23	26	36	31	142
秩和检验		6	16	10	9	16	57
相关分析		12	14	7	8	14	55

7.2 试验设计和资料收集存在的问题

7.2.1 设计不全面, 资料收集混乱, 分析不清楚

【案例7-1】 原文目的探讨细胞病理学检查在经支气管镜超声引导针吸活检术(EBUS-TBNA)诊断中的作用及用于表皮生长因子受体(EGFR)基因突变分析的可行性。原作者收集某医院EBUS-TBNA的连续病例200例, 其中140例进行了穿刺组织条病理学检查和(或)纵隔镜活检和(或)外科手术切除(称为组织学对照)。细胞学/组织学诊断阴性计为阴性, 可疑和阳性计为阳性。结果200例患者中细胞学评价为标本不充足, 无法诊断有24例, 取样成功176例。细胞学定性诊断的准确率(表7-3): 细胞学诊断为良性病变42例, 可疑为癌12例, 恶性肿瘤122例。与组织学诊断(140例)对照, EBUS-TBNA细胞学诊断的敏感度为91.3%, 特异度为93.8%, 总准确率为92.1%; 与组织学和临床共同诊断(200例)对照, 细胞学诊断的敏感度为94.4%, 特异度为100.0%, 总准确率为96.0%。

表7-3 EBUS-TBNA细胞学诊断与组织学及组织学和临床共同诊断比较

细 胞 学	组织学诊断例数			组织学和临床共同诊断例数		
	阳 性	阴 性	合 计	阳 性	阴 性	合 计
阳性	84	3	87	134	0	134
阴性	8	45	53	8	58	66
合计	92	48	140	142	58	200

【案例辨析】 原文作者为了解细胞病理学检查在经支气管镜超声引导针吸活检术(EBUS-TBNA)诊断中的作用, 回顾性调查进行该活检术患者200例, 其中176例患者的标本细胞学可以诊断, 也就是只有176例患者的标本经细胞学诊断是成功的, 但在表7-3中可见, 经细胞学诊断的标本200例, 所以对不成功的24例标本的结果也进行了分析。那么, 这样的分析结果可信吗? 另外, 原作者在细胞学诊断与组织学和临床共同诊断结果比较分析时, 把没有组织学诊断只有临床诊断结果的病例也包含在组织学和临床共同诊断的结果中, 其中共同诊断结果判定条件是什么? 原文没有任何说明, 这给读者会带来很大的困扰, 也不利于对结果的正确理解和判定。

原文作者表述在200例行活检术(EBUS-TBNA)患者中140例有组织学诊断, 见表7-3, 140例有组织学诊断的患者, 其活检标本都符合细胞学评价, 但是在原文资料收集和结果中原作者没有任何说明。

【案例释疑】 原文作者的研究目的很明确, 想了解细胞病理学检查在经支气管镜超声引导针吸活检术(EBUS-TBNA)诊断中的作用。那么, 对于收集的200例患者中176例标本合格, 由此可以计算该活检术(EBUS-TBNA)的采样成功率, 对于24例不合格标本的患者不应进入到下面的分析之中; 在此基础上, 明确176例受试对象中有多少个体进行了穿刺组织条病理学检查和(或)纵隔镜活检和(或)外科手术切除(即有组织学对照); 然后在这个样本量的基础上进行细胞学诊断与组织学诊断对照分析; 对于与组织学和临床共同诊断的对照分析, 应该也在两种检测都有的受试对象中进行, 能够保证同质性, 如果想扩大样本量, 也应该在文章的“材料和方法”中阐述共同诊断的标准和原则。也就是说, 本例是对176例细胞学诊断合格的患者中的140例标本又进行了组织学诊断, 同时, 还对这176例患者进行了临床观察。

7.2.2 受试对象少, 样本代表性差

【案例7-2】 原文目的探讨胰腺实性假乳头状肿瘤(SPT)临床病理学特征和恶性转移潜能相关性。收集45例SPT患者的临床随访资料, 分为非转移组(41例)和转移组(4例), 将两组的临床特征(年龄、

性别、肿瘤部位、肿瘤大小、肿瘤包膜是否存在)、形态学特点(细胞的异型性、核的分级、核分裂象、血管瘤栓、坏死等)及免疫组织化学结果同临床随访资料进行综合比较分析。结果发现非转移组和转移组的年龄、性别、肿瘤部位、肿瘤平均直径和大体包膜情况比较,没有明显差异。单独形态参数(核分级、核分裂象、细胞异型性、血管内瘤栓、渐进性坏死巢、凝固性坏死)并不能区分非转移组和转移组,但核分裂象、血管内瘤栓和渐进性坏死巢在转移组较非转移组更为常见。依据形态学综合评分,非转移组 41 例中,39 例(95.1%)综合得分 <5 分,转移组综合得分全部 ≥5 分。免疫组织化学非转移组与转移组比较,β-catenin、p53、cyclin D1、CD34 差异无统计学意义;核增殖指数 Ki-67 差异有统计学意义。由此,原作者得出结论:SPT 的核分裂象、血管内瘤栓和渐进性坏死巢对于预测肿瘤的恶性转移潜能具有一定的参考意义;形态学综合评分对于预测 SPT 的恶性转移潜能具有重要价值。

【案例辨析】 原文作者目的:分析胰腺实性假乳头状肿瘤(SPT)的临床病理学特征和恶性转移的相关性,收集到的 45 例 SPT 患者中,仅有恶性转移患者 4 例,故转移组样本量为 4,非转移组样本量为 41,进行了临床特征、形态学特点等多种指标比较。用 4 例恶性转移患者的特征来代表所有恶性转移患者的特征,其代表性较差;另外,41 例的非转移组与 4 例的转移组进行比较,样本量相差也很悬殊,这样的结果也不会令人信服。

【案例释疑】 对于任何一项研究在设计阶段都应该计算所需要的样本量,以满足重复原则和保证研究结果的效能。因此,在资料收集时出现了样本量少的问题,应延长研究的时间以保证有更多的受试对象入选,以便最终获得充足且合格的样本量。但如果该疾病恶性转移率较低,无法收集到较多的恶性转移患者,也可以考虑采用 1:M 匹配,这样可以增加分析结果的可信度,降低由于样本量小而带来的代表性差、可信度低的问题。

7.3 定量资料统计分析中存在的问题

【案例 7-3】 原文目的探讨 TNM 病理分期 T1 (pathologic T1, pT1) 期肺腺癌中间质浸润分级对患者预后影响。T1 期肺腺癌 85 例,根据间质浸润在肿瘤中的部位将每例肿瘤中间质浸润的程度分为 1~3 级,1 级:17 例(20%),2 级:12 例(14%),3 级:56 例(66%),间质浸润各级别病例的临床病理学特征主要测量了年龄、性别、吸烟史、肿瘤大小、淋巴血管侵犯率、淋巴结转移率、病理学分期。原作者采用 *t* 检验分析不同间质浸润分级的肺腺癌患者年龄和肿瘤大小。结果认为:①年龄在各级病例间的差异均无统计学意义;②肿瘤大小:除 1 级病例小于 3 级病例(*t* 检验, *P* = 0.005)外其余各级病例间的差异无统计学意义。资料见表 7-4。

表 7-4 pT 期肺腺癌间质浸润各级别病例的临床病理学特征

项 目		各间质浸润分级的例数		
		1 级	2 级	3 级
年龄(岁)	范围	41~76	48~79	28~81
	平均数	61.56	61.17	58.39
肿瘤最大直径(cm)	范围	0.5~3.0	1.0~3.0	1.0~3.0
	平均数	1.87	2.23	2.38

【案例辨析】 从原文可以看出,年龄和肿瘤大小都属于定量资料,而且间质浸润的程度 1 级、2 级、3 级是三组间的比较,但在统计学分析和结果阐述部分对资料分析只叙述采用 *t* 检验,但 *t* 检验方法只能用于不多于两组的资料间的比较,三组间比较并不适用 *t* 检验。如果三组或三组以上的定量资料比较采用 *t* 检验,其后果会导致累计 I 类错误概率增大。

【案例释疑】三组或三组以上定量资料比较，通称为定量资料多组比较。若属于单因素设计范畴，一般采用单因素多水平设计方差分析，或 Kruskal-Wallis H 检验。前者也称为单因素方差分析(One-way ANOVA)，属于参数检验，要求资料满足独立性、正态性和方差齐性，如果三组资料的方差分析结果 $P < 0.05$ ，则认为三组之间至少有两组的差异有统计学意义，应该进一步进行多重比较，通常采用的方法有 LSD 法、SNK 法、Dunnnett 检验、Duncan 法等；Kruskal-Wallis H 检验为非参数检验，不需要资料服从一个已知分布，适用范围较广，其结果同方差分析类似，若 $P < 0.05$ ，则认为三组之间至少有两组的差异有统计学意义，应该进一步进行多重比较。

7.4 定性资料统计分析中存在的问题

7.4.1 误用一般四格表资料 χ^2 检验分析配对设计四格表资料

【案例7-4】原文目的分析比较等位基因特异性寡核苷酸聚合酶链反应(ASO-PCR)和双环探针特异引物荧光 PCR (BPSP-qPCR)的表皮生长因子受体(EGFR)基因突变检测敏感性。原作者收集行 ASO-PCR 检测的非小细胞肺癌标本 96 例，其中 39 例再行 BPSP-qPCR 检测。统计学分析采用 χ^2 检验或 Fisher 四格表确切概率法。结果显示，39 例患者中 BPSP-qPCR 法检测阳性例数为 26 例，阳性率为 66.7%，显著高于 ASO-PCR 法的 16 例($P = 0.02$)，而 BPSP-qPCR 检测阴性的病例中，ASO-PCR 法未检测出阳性。ASO-PCR 法检测阴性的患者中，有 10 例被 BPSP-qPCR 检出有其他部位的突变。

【案例辨析】从原文中可见，有 39 份非小细胞肺癌标本都进行了 ASO-PCR 和 BPSP-qPCR 检测，每个标本都有两个检测结果，应为配对设计。但原作者在文章的“材料和方法”或“结果”等部分没有提到在两个检测方法比较时是按照配对设计进行的，只是在统计分析部分笼统提出定性资料采用 χ^2 检验或 Fisher 四格表确切概率法，而从原文所提供的结果数据来看，原作者采用的分析方法是一般四格表资料 χ^2 检验，最后假设检验的结果是 $P = 0.02$ 。

【案例释疑】该研究中 39 份非小细胞肺癌标本分别进行了 ASO-PCR 和 BPSP-qPCR 检测，原文作者欲分析 ASO-PCR 和 BPSP-qPCR 两种检测方法在 39 份非小细胞肺癌标本中的表达是否有差异？同一个样品用两种不同方法检测，应该采用配对设计四格表资料 χ^2 检验。原文结果修改后见表 7-5(读者可进一步修改此表)，采用配对设计四格表资料 McNemar χ^2 检验(检验结果不一致的频数之间的差别是否具有统计学意义)结果为 $P = 0.002$ ，说明 BPSP-qPCR 法的阳性率高于 ASO-PCR 法。

【案例7-5】原文目的探讨 miR-23a(Homo sapiens miR-23a)与转移抑制因子 1(metastasis suppressor 1, MTSS1)在结肠癌中的表达情况，收集结肠癌患者手术标本 92 例，分别运用原位杂交、免疫组织化学 EliVision 法检测结肠癌组织及癌旁组织中 miR-23a、MTSS1 的表达水平，计数资料(注：正确的叫法为定性资料)采用 χ^2 检验分析。分析结果：92 例结肠癌组织中 miR-23a 阳性表达 80 例，阳性率为 87.0%，92 例癌旁对照组织中 miR-23a 阳性表达 28 例，阳性率为 30.4%，二者相比差异有统计学意义；92 例结肠癌组织中 MTSS1 阳性表达 16 例，阳性率为 17.4%，92 例癌旁对照组织中 MTSS1 阳性表达 76 例，阳性率为 82.61%，二者相比差异有统计学意义($P < 0.01$)。

【案例辨析】从原文中可见，原作者对每名结肠癌患者的癌组织和癌旁组织都进行了 miR-23a 和 MTSS1 两种指标的检测。所以，每个指标都有癌组织和癌旁组织两个检测结果，且来源于同一个个体，所以应该为配对设计。但原作者在文章的“材料和方法”或“结果”等部分没有提到在两

个部位的检测结果为配对设计，只是在统计分析部分笼统提出计数资料(注：正确的叫法为定性资料)采用 χ^2 检验分析，而从原文所提供的结果数据来看，原作者采用的分析方法是一般四格表资料 χ^2 检验。

【案例释疑】 该研究检测了每名患者癌组织和癌旁组织中 miR-23a 和 MTSS1 的表达，原作者欲分析癌组织和癌旁组织中 miR-23a 和 MTSS1 的表达是否有差异？同一个个体两个不同部位的检测结果的比较，应该采用配对设计四格表资料 χ^2 检验。这在统计分析中应该有明确的表述，其结果的表述形式最好为“miR-23a 在结肠癌组织中表达阳性率为 87.0%，在癌旁组织中表达阳性率为 30.4%，在两种组织中表达均为阳性是多少例(%)”，这样的表述结果更能体现该指标在两种组织的表达相同与否。也可以采用表格的形式表达，参见表 7-6(读者可进一步修改此表)。

表 7-5 39 份非小细胞肺癌标本两种方法检测结果

ASO-PC	BPSP-qPCR		合 计
	+	-	
+	16	0	16
-	10	13	23
合计	26	13	39

表 7-6 miR-23a 和 MTSS1 在癌组织和癌旁组织的表达(参考表)

	癌 组 织	癌 旁 组 织		合 计
		+	-	
miR-23a	+			
	-			
	合计			
MTSS1	+			
	-			
	合计			

7.4.2 误用多次一般四格表资料 χ^2 检验分析多组比较定性资料

【案例 7-6】 原文目的探讨 TNM 病理分期 T1 (pathologic · T1, pT1) 期肺腺癌中间质浸润分级对患者预后影响。T1 期肺腺癌 85 例，根据间质浸润在肿瘤中的部位将每例肿瘤中间质浸润的程度分为 1~3 级，1 级：17 例(20%)，2 级：12 例(14%)，3 级：56 例(66%)，间质浸润各级别病例的临床病理学特征主要测量了年龄、性别、吸烟史、肿瘤大小、淋巴血管侵犯率、淋巴结转移率、病理学分期。该作者采用 χ^2 检验分析不同间质浸润分级的肺腺癌患者性别、吸烟史、淋巴血管侵犯率、淋巴结转移率、病理学分期。结果认为：①性别和吸烟史：各级病例间的差异均无统计学意义；②淋巴血管侵犯率：1 级病例为 0，2 级病例为 8%，3 级病例为 32%，经分析 1 级病例低于 3 级病例(χ^2 检验， $P=0.018$)，其余 1 级与 2 级病例间及 2 级与 3 级病例间的差异无统计学意义；③淋巴结转移率：1 级、2 级病例均为 0，3 级病例为 38%，经分析 1 级(χ^2 检验， $P=0.007$)及 2 级病例(χ^2 检验， $P=0.027$)均低于 3 级病例；④病理学分期：1 级、2 级病例均为 IA 期，3 级病例 61% 为 IA 期，39% 高于 IA 期，经分析 1 级(χ^2 检验， $P=0.002$)及 2 级病例(χ^2 检验， $P=0.021$)的病理学分期低于 3 级病例。

【案例辨析】 从原文可以看出，性别、吸烟史、淋巴血管侵犯率、淋巴结转移率、病理学分期都属于定性资料，间质浸润的程度 1 级、2 级、3 级是三组间的比较，但在统计学分析和结果阐述部分原作者只叙述采用 χ^2 检验，而从原文叙述来看，例如淋巴结转移率：1 级、2 级病例均为 0，3 级病例为 38%，经分析 1 级(χ^2 检验， $P=0.007$)及 2 级病例(χ^2 检验， $P=0.027$)均低于 3 级病例，原作者是对间质浸润程度不同的三组进行了直接拆分，做了 3 次两组之间的 χ^2 检验，但对拆分后的检验水准也没有进一步说明，而这样的分析方法会导致犯累计 I 类错误概率增大。

【案例释疑】 三组或三组以上的定性资料，一般称为行 \times 列表资料或 $R \times C$ 列联表资料，对其

分析首先应进行行×列表资料 χ^2 检验,若 $P < 0.05$,则拒绝无效假设 H_0 ,认为多组中至少有两组之间的差异有统计学意义,但这也只能作出总的结论,不能认为任意两组之间差异都有统计学意义。若想知道哪两组之间差别有统计学意义或无统计学意义,需要进行两两比较,一般采用调整检验水准后进行两两比较。该方法在进行两两比较时不能再采用原来的检验水准 $\alpha = 0.05$ 作为是否拒绝 H_0 的标准,因为重复多次的假设检验,将使 I 类错误 α 扩大,必须重新规定检验水准作为拒绝 H_0 的标准。新的检验水准通常需要将欲进行两两比较的次数考虑在内,例如,如果三组定性资料两两比较,需要进行3次,则新的检验水准为:

$$\alpha' = \frac{0.05}{3} = 0.017$$

7.5 等级资料统计分析中存在的问题

7.5.1 误用 χ^2 检验分析等级资料

【案例7-7】 原文目的研究趋化因子受体 CXCR3 在乳腺癌中的表达,分析其与乳腺癌患者相关临床病理指标之间的关系。原作者应用免疫组织化学 SP 法检测 CXCR3 在 80 例乳腺癌、20 例乳腺纤维腺瘤和 15 例正常乳腺组织中的表达,结果根据着色强度乘以阳性细胞比例划分为:0 分为阴性(-),1~3 分为弱阳性(1+),4~6 分为中度阳性(2+),8~12 分为强阳性(3+)。采用 Spearman 秩相关分析 CXCR3 的表达与 80 例浸润性乳腺癌患者各临床病理特征之间的关系;采用 χ^2 检验比较淋巴结转移阴性组和阳性组患者的 CXCR3 的表达,资料见表 7-7。

表 7-7 CXCR3 表达与各临床病理指标的相关性

项 目	例 数	CXCR3 表达(例)				r_s	P 值
		-	1+	2+	3+		
淋巴结状态						0.417	0.000
0	34	6	11	13	4		
1~3	27	1	5	11	10		
4~9	11	0	3	5	3		
>9	8	0	0	2	6		
肿瘤大小(cm)						0.319	0.004
≤2	39	5	16	9	9		
2~5	38	3	8	13	14		
>5	3	0	0	2	1		
...

【案例辨析】 原文材料与方法中描述: CXCR3 表达按照标准划分为:0 分为阴性(-),1~3 分为弱阳性(1+),4~6 分为中度阳性(2+),8~12 分为强阳性(3+),由此可见,该研究是按照多项有序分类资料(等级资料)收集的,从表 7-7 所示的结果也可以进一步验证。在行×列表资料中,如果资料的结果变量是有序的, χ^2 检验只是说明各组在各等级上的频数分布是否相同,无法确定各组总体效应之间差异是否有统计学意义。该文在对淋巴结转移阴性组和阳性组患者的 CXCR3 的表达差异是否有统计学意义时采用了 χ^2 检验,并得出:淋巴结转移阳性组 CXCR3 的表达显著高于淋巴结转移阴性组,差异有统计学意义($P < 0.05$),这样的结果只能说明淋巴结转移阴性组和阳性组患者的 CXCR3 的表达四个水平上的频数分布不同,不能直接得到淋巴结转移阴性组和阳性组患者的 CXCR3 的表达差异有统计学意义。

另外,该研究结果表 7-7 中只列出了各临床病理特征与 CXCR3 表达之间的相关关系,没有说明临床病理特征不同水平的患者之间 CXCR3 表达差异是否有统计学意义,而这种差异的分析结果与相关的分析结果并不是完全一致的。

【案例释疑】 CXCR3 表达为多项有序分类资料(等级资料),分析有序分类资料的效应差异时,应采用秩和检验或 Ridit 分析。所以,对于淋巴结转移阴性组和阳性组患者的 CXCR3 的表达比较可以采用 Wilcoxon 秩和检验。另外,在表 7-7 中,淋巴结状态的分析结果 P 值为 0.000,这种写法是错误的,即使 P 值很小,也不会为 0;如果 P 值比较小,可以表达为 $P < 0.0001$ 。

7.5.2 误用多次两样本秩和检验分析单因素多水平设计有序资料

【案例 7-8】 原文目的观察上皮间质转化因子 Twist 转录因子和相关蛋白 E-和 N-钙黏蛋白在乳腺癌中的表达及其与患者临床病理指标的关系。采用免疫组织化学 SP 法检测 56 例浸润性导管癌(IDC)、38 例浸润性小叶癌(ILC)和 41 例导管内癌以及 10 例正常乳腺组织中 Twist、E-和 N-cadherin,表达强度的判定标准: Twist、N-cadherin 阳性肿瘤细胞数 $\leq 20\%$ 者均视为阴性;阳性肿瘤细胞数 $> 20\%$ 视为阳性高表达; E-cadherin 与正常组织中的阳性对照相比,阳性表达范围低于正常的 10% 者视为阴性,阳性表达范围是正常的 90% 以上者视为阳性。采用两个独立样本秩和检验分析不同组中三种蛋白的表达。结果表明: ILC 中 Twist 阳性率显著高于 IDC 和导管内癌($P = 0.002$, $P = 0.000$); IDC 和导管内癌中 E-cadherin 阳性率显著高于 ILC(均 $P = 0.000$); IDC 和 ILC 中 N-cadherin 阳性率显著高于导管内癌($P = 0.033$, $P = 0.001$)。

【案例辨析】 原文中结果部分可见,原作者比较了 IDC、ILC 和导管内癌三组中三种蛋白指标表达,采用的是两个独立样本比较秩和检验,直接把三组中的每两组进行了一次两个独立样本秩和检验,这样进行了 3 次假设检验,但对于检验水准没有进一步的说明,这样的分析方法会导致累计 I 类错误概率增大。另外,原作者在原文中以 $P = 0.000$ 来表示,这样的表示方法是错误的,只有不可能发生的事件的概率才为 0。因此,无论假设检验获得的 P 值有多小,也不可能为 0;如果结果很小,统计软件结果虽表示为 0.000,一般在文章中应表示为 $P < 0.0001$ 。

【案例释疑】 原文中三种蛋白指标按照等级资料收集,将三组不同类型乳腺癌比较应该采用 Kruskal-Wallis H 检验,由此得到的结果为三组总体比较差异有无统计学意义;若 $P < 0.05$,差异有统计学意义,应该进行三组间两两比较,可以采用 Wilcoxon 秩和检验,但检验水准要求进行分割,即新的检验水准通常是把 0.05 被检验的次数分割,例如,如果三组比较,需要进行 3 次两两比较,则新的检验水准为 $\alpha' = \frac{0.05}{3} = 0.017$,将这个检验水准作为两两比较假设检验的判断标准。

7.5.3 误用 Mann-Whitney U 检验分析配对设计有序资料

【案例 7-9】 原文目的对比上皮性原发癌、转移癌与其癌栓的组织形态学及 E-cadherin 和波形蛋白表达情况,收集 68 个同时具有癌组织(包含原发癌或淋巴结转移癌)及癌栓的石蜡组织标本,采用免疫组织化学(SP 法)检测 E-cadherin 和波形蛋白表达。Mann-Whitney 检验分析结果显示,癌灶组织与其癌栓的 E-cadherin 和波形蛋白表达差异均无统计学意义($Z = -0.248$, $P = 0.804$; $Z = -0.199$, $P = 0.842$)。

【案例辨析】 Mann-Whitney 检验即为 Mann-Whitney U 检验,经常用于两独立样本比较的非参数检验。该文比较癌灶组织和癌栓中的 E-cadherin 和波形蛋白表达是否有差异,为典型的配对设计定性资料,不应该使用 Mann-Whitney U 检验。

【案例释疑】 由于该文中癌灶组织和癌栓为同一个个体的两个不同部位,应该属于配对设计;

而检测指标 E-cadherin 和波形蛋白表达都是按照等级资料收集的, 所以应该使用配对设计等级资料比较的分析方法, 常使用的是 Wilcoxon 符号秩检验。

7.5.4 误用 Wilcoxon 秩和检验分析配对设计有序资料

【案例 7-10】 原文目的探讨乳腺癌新辅助化疗前核芯针穿刺活检 (CNB) 标本和化疗后手术切除标本分子标志物表达。收集 209 例接受新辅助化疗的乳腺癌患者 CNB 和手术切除标本, 检测分子标志物雌激素受体 (ER)、孕激素受体 (PR) 和 HER2 的表达信息。ER 和 PR 结果判定标准: ①10% 以上的细胞核阳性判定为阳性; ②依染色强度分成弱阳性 (1+)、中等阳性 (2+) 和强阳性 (3+)。HER2 结果判定标准: ①0 分: 浸润性癌细胞未见着色; 1+: 任何比例浸润性癌细胞呈现弱的、不完整的细胞膜着色, 或不足 10% 的浸润性癌细胞呈现弱的、完整的细胞膜着色; 2+: 10% 以上浸润性癌细胞呈现不均匀的或弱的完整细胞膜着色, 或 30% 及以下的浸润性癌细胞呈现完整的细胞膜强阳性着色; 3+: 30% 以上浸润性癌细胞呈均匀的细胞膜强阳性着色。应用 Wilcoxon 秩和检验分析新辅助化疗前后肿瘤组织分子标志物的表达是否存在改变。结果显示治疗前后仅 HER2 表达改变的差异存在统计学意义 ($P=0.049$)。

【案例辨析】 从原文可以看出, 原作者收集了接受新辅助化疗的乳腺癌患者 209 例, 新辅助化疗前均进行 CNB, 化疗结束后 4 周内实施乳腺癌根治术。收集患者 CNB 和手术切除标本, 检测分子标志物雌激素受体 (ER)、孕激素受体 (PR) 和 HER2 的表达信息。可见, 无论原作者检测哪一个指标, 对于一个患者来说, 都会有来自不同标本的两个结果, 即获得的资料应该为配对设计定性资料。但原作者在文中采用 Wilcoxon 秩和检验来分析新辅助化疗前后肿瘤组织分子标志物的表达是否有改变, 明显地将资料看作为来自成组设计定性资料。这样的分析结果不能令人信服。

【案例释疑】 由于该文中分子标志物雌激素受体 (ER)、孕激素受体 (PR) 和 HER2 在患者的 CNB 和手术切除标本中都进行检测, 即为同一个指标在两个不同部位的分布, 应该属于配对设计; 而检测指标的表达都是按照有序资料收集的, 所以应该使用配对设计有序资料的分析方法, 经常使用的是 Wilcoxon 符号秩检验。

7.6 预后分析常见错误

7.6.1 分析结果意义不明确

【案例 7-11】 原文目的探讨 TNM 病理分期 T1 (pathologic · T1, pT1) 期肺腺癌中间质浸润分级对患者预后影响。T1 期肺腺癌 85 例, 根据间质浸润在肿瘤中的部位将每例肿瘤中间质浸润的程度分为 1~3 级, 临床病理学特征主要测量了年龄、性别、吸烟史、肿瘤大小、淋巴血管侵犯率、淋巴结转移率、病理学分期。各因素对预后影响进行了单因素分析, 显示间质浸润分级、病理学分期、淋巴血管侵犯及淋巴结转移与预后相关。因此, 对这些与预后相关的因素进行 Cox 多因素回归分析, 结果为病理学分期为独立预后因素, 间质浸润分级、淋巴结转移及淋巴血管侵犯不是独立预后因素。资料见表 7-8。

表 7-8 pT1 期肺腺癌间质浸润分级、淋巴结转移、淋巴血管侵犯及
病理学分期与预后相关性 Cox 多因素回归分析

检测项目	风险率	95% 可信区间	P 值
间质浸润分级 (3 级)	3.345	0.990 ~ 11.300	0.052
淋巴结转移 (有)	2.003	0.485 ~ 8.263	0.337

续表			
检测项目	风险率	95%可信区间	P值
淋巴血管侵犯(有)	0.426	0.164 ~ 1.103	0.079
病理学分期(有)	7.239	2.827 ~ 18.535	<0.001

【案例辨析】 表7-8为原文所示的Cox回归分析结果，其只列出3项结果，没有提供该分析方法其他必要的结果，结果表达极其不规范，不能让读者正确判断和验证该分析结果正确与否。该结果中也列出了无统计学意义的因素，如间质浸润分级、淋巴结转移及淋巴血管侵犯，而一般情况下，无统计学意义的因素在经过变量筛选后都不会保留在最终的结果表中。由此可见，原作者在进行Cox回归分析时未选用合适的变量筛选方法，导致所有的变量都被保留在回归方程中。众所周知，因素之间总会是相互影响和制约的，由于某些无统计学意义的因素存在可能使得对结果有影响的因素变为无影响或影响结果发生程度上的改变，所以，自变量筛选时应选择合适的方法。

【案例释疑】 Cox回归分析为多因素分析方法，一般变量筛选方法多采用前进法或后退法，而不采用全模型的分析方法。该研究应该删除对预后影响无统计学意义的自变量，而结果表示一般包含如下指标：回归系数估计值 β 及其标准误差、对回归系数进行假设检验的Wald值及其对应的具体P值、RR及其95%CI(注：CI为置信区间的缩写)。

7.6.2 结果描述过于简单，表述不清

【案例7-12】 原文目的探讨原发乳腺淋巴瘤(PBL)的临床病理、免疫组织化学特征及其与预后的关系。作者按Ann Arbor临床分期标准及Wiseman和Liao的诊断标准对40例PBL患者进行分期，按照WHO(2008年)造血和淋巴瘤系统肿瘤分类分型，并进行了随访。采用Cox比例风险模型进行随访资料预后分析，发现LDH水平、肿瘤数目、IPI指数影响患者的总生存率(P值分别为0.014、0.004、0.044)，这3个因素同样也影响患者的5年无病生存率(P值分别为0.029、0.015、0.011)。

【案例辨析】 原文作者采用Cox比例风险模型对随访资料进行预后分析，在结果中，原作者只给出影响因素的P值，其他任何Cox比例风险模型分析所给出的必要结果都没有列出，结果表示极其不规范，不能让读者正确判断和验证该分析结果正确与否，也难于了解各自变量对预后影响的风险率是如何的。

【案例释疑】 对于Cox比例风险模型分析，一般要提供如下结果：回归系数估计值 b 及其标准误差、对回归系数进行假设检验的Wald值及其对应的具体P值、RR及其95%CI。即使由于文章篇幅所限，不能列出表格，但一般我们更倾向于在结果中表述RR(95%CI)，这样的结果不仅能表示影响因素是否有统计学意义，同时也能提供影响因素的效应大小和方向。如若列表表示，可以参照表7-9。

表 7-9 Cox 比例风险模型分析结果

变 量	b_i (估计值)	S_{b_i} (标准误差)	Wald χ^2	P 值	RR	RR 的 95%置信区间	
						下 限	上 限
X_1							
X_2							
...							

7.7 资料表达与结果描述中存在的统计学错误

7.7.1 错误定义检验水准

【案例7-13】 原文在统计学分析中表述如下：采用SPSS 18.0软件统计，分类资料(年

龄、性别、吸烟史、病理类型和突变结果)采用卡方检验或 Fisher 四格表确切概率法,检验效能 $\alpha = 0.05$ 。

【案例辨析与释疑】原文提出 $\alpha = 0.05$ 作为检验效能,实际 α 是检验水准的表示符号,一般设定为 0.05;而检验效能也不可能为 0.05,一般要达到 80% ~ 90% 才有意义。

【案例 7-14】原文在统计学分析中表述如下:全部分组标志定量数据以“均数 \pm 标准差”表示。采用双侧 t 检验比较各组平均值,可信度为 95%。

【案例辨析与释疑】原文作者在统计学分析部中提出“可信度为 95%”,欲表达他在假设检验中使用的判断标准,即检验水准 α 。通常情况下对检验水准 α 的说明,不使用“可信度为 95%”的形式,容易误解,而是表述为“以 $P < 0.05$ 为差异有统计学意义”。

7.7.2 统计量标识不准确

【案例 7-15】表 7-10 为某病预后的 logistic 回归分析结果,采用 u 值作为统计量来确定各指标的 P 值。

【案例辨析与释疑】logistic 回归分析结果中采用 u 值作为统计量,虽然对结果没有任何影响,但一般情况下 logistic 回归的统计量常常使用 Wald χ^2 值,大多数软件(如 SAS、SPSS、等)也采用 Wald χ^2 值作为 logistic 回归分析中对回归参数进行假设检验的统计量。因此,在描述 logistic 回归分析的结果时,建议使用 Wald χ^2 值,让读者易于理解,不会产生误解。

表 7-10 影响预后的多因素 logistic 逐步回归分析

参 数 名	估 计 值	标 准 误 差	u 值	P 值	RR	RR 的 95%CI
常数项	-0.928	0.572	1.622	0.1049		
AC	-1.505	0.604	2.494	0.0126	0.22	0.068 ~ 0.724
SCC	-2.179	0.669	3.257	0.0011	0.11	0.030 ~ 0.420
...

7.7.3 P 值表达错误

【案例 7-16】这是一篇文章的结果表示:ILC 中 Twist 阳性率显著高于 IDC 和导管内癌($P = 0.002$, $P = 0.000$);IDC 和导管内癌中 E-cadherin 阳性率显著高于 ILC(均 $P = 0.000$);IDC 和 ILC 中 N-cadherin 阳性率显著高于导管内癌($P = 0.033$, $P = 0.001$)。

【案例辨析与释疑】原文作者用 $P = 0.000$ 来表示假设检验的结果,这样的表示方法是错误的,只有不可能发生事件的概率才为 0。因此,无论假设检验的 P 有多小,也不可能为 0;如果结果很小,虽然统计软件将结果表示为 0.000,一般在文章中表示为 $P < 0.0001$ 。

7.7.4 相关性分析符号使用混乱

【案例 7-17】原文叙述如下:采用 Pearson 相关系数分析连续变量之间相关性;采用 Spearman 等级相关系数分析等级变量之间相关性。结果为: NKG2D 的表达不仅与骨肉瘤的分化呈明显负相关($r = -0.552$, $P = 0.040$),而且与临床分期也呈明显负相关($r = -0.592$, $P = 0.026$)。

【案例辨析与释疑】原文作者用 r 来表示相关系数,但没有区分 r 代表的是 Pearson 相关系数还是 Spearman 等级相关系数?一般来说, r 代表的是 Pearson 相关系数, r_s 代表 Spearman 等级相关系数。从原文叙述可见,骨肉瘤的分化程度和临床分期都属于等级变量,应该使用 Spearman 等级

相关系数来分析，但原文结果均用 r 来表示结果，容易让读者误认为都采用 Pearson 相关分析。因此，如果采用 Spearman 等级相关系数最好表示为 r_s 。

7.7.5 图表错误

【案例 7-18】 这是在一篇文章中出现两条生存曲线，探讨 TNM 病理分期 T1 期肺腺癌中间质浸润分级对患者预后影响。图 7-1 为 85 例肺腺癌患者生存曲线，图 7-2 为间质浸润 3 个分级的患者生存曲线。85 例肺腺癌患者按照间质浸润程度分为三级：1 级 17 名患者，2 级 12 名患者，3 级 56 名患者。

【案例辨析与释疑】 图 7-1 和图 7-2 中所描述的资料为同一个样本，只是从不同的角度来阐述预后生存情况。由图 7-2 可见，该组资料中存在有截尾数据，但是在图 7-1 中没有看到有截尾数据的标识，容易给人造成误解，以为不是同一批资料。另外，两幅图的纵坐标变量名称不同，应仔细斟酌。

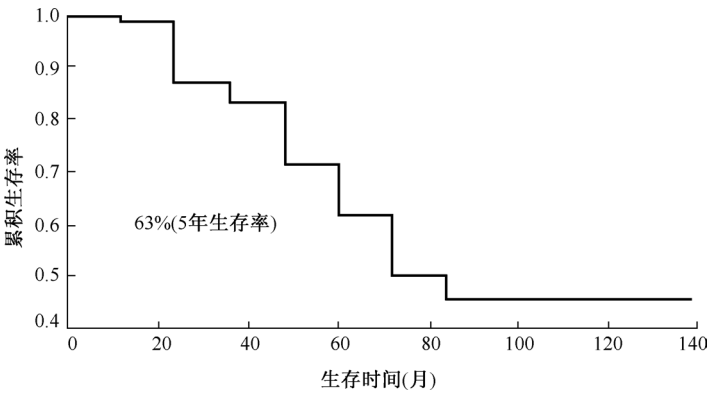


图 7-1 85 例肺腺癌患者生存曲线

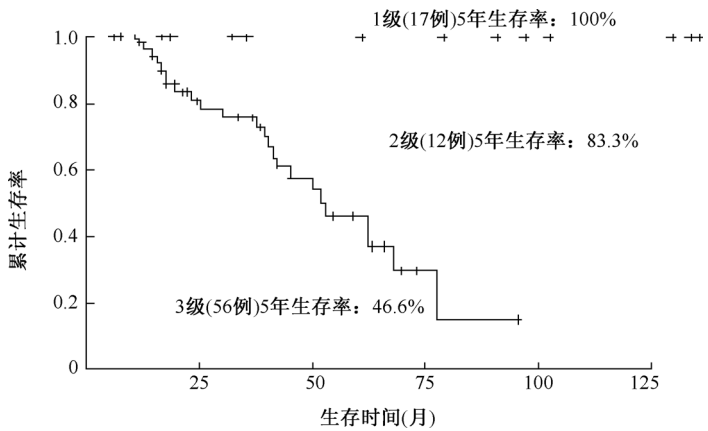


图 7-2 间质浸润 3 个分级的患者生存曲线

【案例 7-19】 这是在一篇文章中出现两条生存曲线，探讨原发乳腺淋巴瘤 (PBL) 的临床病理、免疫组织化学特征对患者预后影响。图 7-3 为弥漫性大 B 细胞淋巴瘤 (DLBCL) 患者生存曲线，图 7-4 为弥漫性大 B 细胞淋巴瘤 (DLBCL) 患者分为 GCB 亚型和非 GCB 亚型的生存曲线。

【案例辨析与释疑】 由图 7-4 可见，该组资料中存在有截尾数据，但是在图 7-3 中没有看到有截尾数据的标识，容易给人造成误解，以为不是同一批资料。

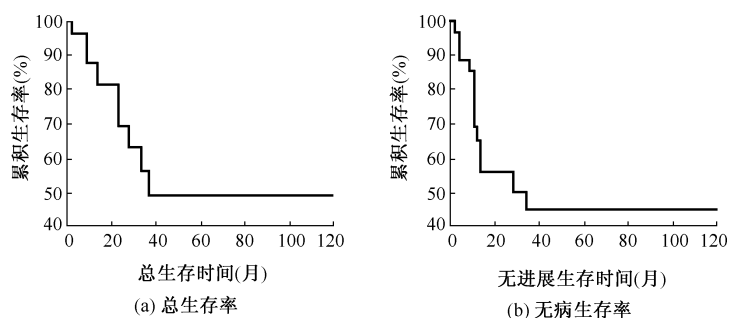


图 7-3 弥漫性大 B 细胞淋巴瘤 (DLBCL) 患者生存曲线

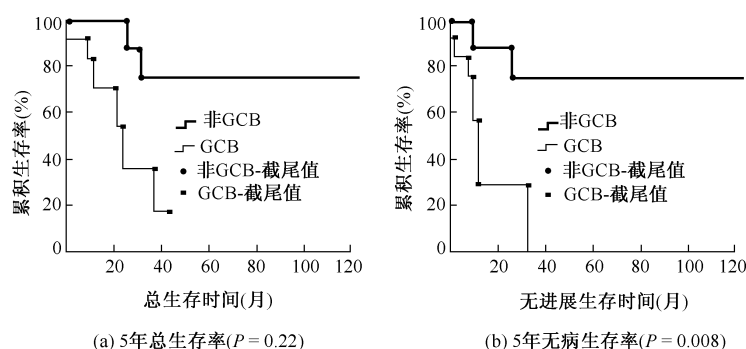


图 7-4 弥漫性大 B 细胞淋巴瘤 (DLBCL) 患者分为 GCB 亚型和非 GCB 亚型的生存曲线

7.8 统计学分析方法小结和 SPSS 菜单驱动的实现

统计学分析方法一般按照描述统计和推断统计来划分,不同资料类型的描述统计和推断统计分析方法不同。

7.8.1 定量资料的分析方法

(1) 描述统计。

定量资料常需要描述频数分布特征,即集中趋势和离散程度。对于服从正态或近似正态分布的定量资料,经常采用均数(\bar{x})和标准差(s)来描述;对于偏态分布或分布不明确的定量资料,常采用中位数(M)和四分位间距(Q)或全距(R)来描述;对于服从对数正态分布的定量资料,一般将其进行对数转换后,按照正态或近似正态分布的定量资料来处理。

(2) 假设检验。

假设检验是推断统计最常使用的方法。定量资料常用的有 t 检验和方差分析 (ANOVA),前者用于单因素两水平设计定量资料两个总体均数的比较,后者用于单因素多水平设计或某种多因素设计定量资料三个或三个以上总体均数的比较,但都要求资料服从正态分布且方差齐。对于不服从正态分布或方差不齐的定量资料应该使用非参数检验(参考等级资料的分析方法)。

7.8.2 定性资料的分析方法

(1) 描述统计。

为便于组间定性资料的比较,经常采用相对数来描述定性资料。常用的相对数有率(Rate)、构

成比 (Constituent Ratio) 与相对比 (Relative Ratio)。在应用相对数分析资料时, 应该注意计算相对数的分母不宜太小, 同时一定不要把构成比当做率来使用。

(2) 假设检验。

常用 χ^2 检验分析定性资料, 包括一般四格表资料 χ^2 检验、配对设计四格表资料 χ^2 检验和行列表资料 χ^2 检验。前二者为两个总体率或构成比的比较, 后者为完全随机设计三个或三个以上总体率或构成比的比较。每种假设检验方法都有最适条件, 要根据资料的特征选择正确的方法。

7.8.3 有序资料的分析方法

(1) 描述统计。

一般情况下, 有序资料常用构成比表述。

(2) 假设检验。

有序资料采用非参数检验来分析, 最常用秩和检验。Wilcoxon 符号秩和检验常用于单组设计和配对设计定性资料差异性分析; Wilcoxon 秩和检验常用于单因素两水平设计定性资料差异性分析; Kruskal-Wallis H 检验常用于单因素多水平设计定性资料差异性分析。

7.8.4 多因素分析

(1) 多重线性回归分析。

用于刻画一个因变量随多个自变量变化而变化的线性关系, 要求因变量 (Y) 为定量资料, 通过偏回归系数和标准化偏回归系数确定各自变量对因变量的影响大小。分析结果表述形式见表 7-11。

表 7-11 多重线性回归分析结果

变 量	b_i (估计值)	S_{b_i} (标准误差)	b_i' (标准化偏回归系数)	t	P
常数项					
X_1					
X_2					
...

(2) Logistic 回归分析。

Logistic 回归分析主要用于筛选疾病的危险因素或预后因素和评价治疗措施的效果, 通常是以疾病的死亡与否、治愈与否等结局为因变量 [二分类的资料, 注意: 真正的因变量为某种结局 (如死亡) 发生的概率], 以疾病发生和预后的影响因素为自变量建立模型。主要利用 OR 值来反映各自变量对因变量的影响程度。分析结果表述形式见表 7-12。

表 7-12 Logistic 回归分析结果

变 量	b_i (估计值)	S_{b_i} (标准误差)	Wald χ^2	P	OR (95% CI)
X_1					
X_2					
...

(3) Cox 比例风险模型。

Cox 比例风险模型是一种允许资料有“删失 (或截尾)”数据存在的, 可以同时分析众多因素对生存时间影响的多变量生存分析方法, 是一种半参数方法, 也简称 Cox 回归模型。其因变量为生存

时间,通常还伴随一个指示变量,标志每位受试者的生存时间是准确数据还是截尾数据,主要利用 RR 值来反映各自变量对因变量的影响。分析结果表述形式见表 7-13。

表 7-13 Cox 比例风险模型分析结果

变 量	b_i (估计值)	S_{b_i} (标准误差)	Wald χ^2	P	RR (95% CI)
常数项					
X_1					
X_2					
...

7.8.5 用 SPSS 软件(菜单驱动法)实现常用统计分析

(1) 正态性检验。

Analyze→Descriptive Statistics→Explore, 打开 Explore 对话框,从源变量中选择数值型变量调入 Dependent List 框;从源变量中选择分组变量(如果有的话)调入 Factor List 框,表示按不同组进行分析;单击 Plots 按钮,打开对话框,选中 Normality plots with test 复选框,单击 Continue 按钮,单击 OK 按钮,输出结果。当正态性检验(test of normality)的 sig(即为 P 值)小于 0.05 时,不能认为总体的分布服从正态分布。

(2) 已知总体与未知总体均数比较 t 检验。

Analyze→Compare Means→One-Sample T Test, 打开 One-Sample T Test 对话框,将分析变量 x 调入 Test Variable(s) 框,将 Test Value 值由 0 改为已知总体均数。单击 OK 按钮,输出结果。

(3) 配对设计一元定量资料均数比较 t 检验。

Analyze→Compare Means→Paired-Samples T Test, 打开 Paired-Samples T Test 对话框,同时选中进行配对的两个变量并调入 Paired Variable 框,单击 OK 按钮,输出结果。

(4) 单因素两水平设计一元定量资料总体均数比较 t 检验。

Analyze→Compare Means→Independent-Samples T Test, 打开 Independent-Samples T Test 对话框,将检测变量调入 Test Variable(s) 框,将分组变量调入 Grouping Variable 框,单击 Define Groups 按钮,打开 Define Groups 对话框,输入分组类别。单击 Continue 按钮,单击 OK 按钮,输出结果。

(5) 单因素多水平设计一元定量资料方差分析。

Analyze→Compare Means→One-Way ANOVA, 打开 One-Way ANOVA 对话框。将检测变量调入 Dependent List 框,将分组变量调入 Factor 框,单击 Continue 按钮,单击 OK 按钮,输出结果。

(6) 单因素两水平或多水平设计二值或多值名义资料的 χ^2 检验。

Analyze→Descriptive Statistics→Crosstabs, 打开 Crosstabs 对话框,将行变量调入 Row(s) 框,列变量调入 Columns(s) 框。单击 Statistics 按钮,选中 Chi-square 按钮。单击 OK 按钮,输出结果。一般四格表资料 χ^2 检验和行列表资料 χ^2 检验都适用该过程。

(7) 配对设计一元定量资料 Wilcoxon 符号秩检验。

Analyze→Nonparametric Test→2 Related Samples, 打开 Two-Related-Samples Tests 对话框。将配随设计的两个变量选中且同时调入 Test Pair(s) List 框,单击 OK 按钮,输出结果。

(8) 单因素两水平设计一元定量资料 Wilcoxon 秩和检验。

Analyze→Nonparametric Test→2 Independent Samples..., 打开 Two-Independent-Samples Tests 对话框。将检测变量调入 Test Variable List 框,将分组变量调入 Grouping Variable 框;单击 Define Groups 按钮,打开“Two-Independent-Samples”: Define Groups”对话框,输入分组类别。单击 Continue 按钮,单击 OK 按钮,输出结果。

(9) 简单线性相关分析。

Analyze→Correlate→Bivariate, 打开 Bivariate Correlation 对话框, 将分析变量都调入 Variables 框。选中 Pearson 选项, 单击 OK 按钮, 输出结果。如果计算 Spearman 等级相关系数, 则在 Bivariate Correlation 对话框中, 选中 Spearman 选项, 单击 OK 按钮, 输出结果。

(10) 线性回归分析。

Analyze→Regression→Linear, 打开 Linear Regression 对话框。将因变量调入 Dependent 框, 自变量调入 Independent 框, 在 Method 框选取合适的变量筛选方法, 单击 OK 按钮, 输出结果。简单线性回归和多重线性回归都适用。

(11) Logistic 回归分析。

Analyze→Regression→Binary Logistic, 打开 Logistic Regression 对话框。将因变量调入 Dependent 框, 自变量调入 Covariates 框, 在 Method 框选取合适的变量筛选方法, 单击 OK 按钮, 输出结果。

(12) Cox 比例风险模型分析。

Analyze→Survival → Cox Regression , 打开 Cox Regression 对话框。将生存时间调入 Time 框, 将结局变量调入 Status 框, 自变量调入 Covariates 框, 在 Method 框选取合适的变量筛选方法, 单击 OK 按钮, 输出结果。

参 考 文 献

- [1] 胡良平. 容易混淆的试验设计类型的辨析, 中国卫生统计, 1993, 10(2): 1-3.
- [2] 方积乾. 生物医学研究的统计方法. 北京: 高等教育出版社, 2010, 15-490.
- [3] 彭敏宁, 昌兰. 医学论文中常见统计学错误案例分析. 国外医学-生理病理科学与临床分册, 2003, 23(6): 664-668.
- [4] 孙振球. 医学统计学(第3版). 北京: 人民卫生出版社, 2010, 11-169.
- [5] 胡良平. 医学期刊中常见的统计学错误及对策. 中国卒中杂志, 2006, 1(12): 865-878.
- [6] 刘桂芬. 医学统计学(第2版). 北京: 中国协和医科大学出版社, 2007, 9-138.
- [7] 胡良平. 统计资料表达与描述方面错误辨析与释疑. 基础医学与临床, 2007, 27(4): 476-480.
- [8] 李洪成. SPSS 18 数据分析基础与实践. 电子工业出版社, 2010.
- [9] 张文彤. SPSS 统计分析教程: 基础篇. 北京: 北京希望电子出版社, 2002, 23-197.
- [10] 陈平雁. 研究生教学用书配套教材·SPSS 13.0 统计软件应用教程. 北京: 人民卫生出版社, 2005: 10-157.
- [11] 何丽娟(译者), 朱红译(译者). SPSS 统计应用与解析(第3版). 电子工业出版社, 2009, 15-229.
- [12] 张文彤, 闫洁. SPSS 统计分析基础教程. 北京: 高等教育出版社, 2004, 10-156.

第 8 章 细胞与分子免疫学统计设计与数据分析

《细胞与分子免疫学杂志》主要收录生命科学领域中有关细胞免疫、抗体工程和分子免疫学等方面的理论及应用研究成果。

本章针对该杂志近 5 年论文中统计设计与统计分析出现的较严重问题进行辨析与释疑。包含了研究设计、定量资料统计分析、定性资料统计分析和多因素分析等方法。其中该杂志统计问题以“用单因素方差分析处理多因素析因设计和重复测量设计定量资料”最为多见，本章也将对该类方法出现的问题重点阐述，为广大细胞与分子免疫学科研工作者提供参考。

8.1 统计设计错误辨析

8.1.1 对照不全和采用的统计学分析方法错误

【案例 8-1】 原文目的观察Ⅱ型胶原蛋白对胶原诱导的关节炎(CIA)大鼠外周血 CD4⁺ CD25⁺ FOXP3⁺ 调节性 T 细胞的影响。方法：将雌性 Wistar 大鼠 40 只随机分为 4 组，即正常对照组、模型对照组、Ⅱ型胶原蛋白治疗组和雷公藤多苷对照组，正常对照组和模型对照组分别用生理盐水灌胃，Ⅱ型胶原蛋白治疗组和雷公藤多苷对照组分别用Ⅱ型胶原蛋白和雷公藤多苷灌胃。观察Ⅱ型胶原蛋白对 CIA 大鼠外周血 CD4⁺ CD25⁺ FOXP3⁺ 调节性 T 细胞的影响。统计方法组间比较采用 *t* 检验，结果见表 8-1 和表 8-2。

表 8-1 各组外周血 CD4⁺ CD25⁺ FOXP3⁺ Treg 水平的比较 (*n*=10, %, $\bar{x} \pm s$)

组 别	CD4 ⁺ CD25 ⁺ T 细胞	CD4 ⁺ FOXP3 ⁺ T 细胞	CD25 ⁺ FOXP3 ⁺ T 细胞
正常对照组	70.1 ± 12.52	93.2 ± 4.13	65.5 ± 12.64
模型组	59.9 ± 13.11	95.4 ± 2.21	49.3 ± 4.20 ^b
CⅡ组	86.7 ± 15.32 ^d	97.1 ± 1.44	84.3 ± 15.81 ^d
TWP 组	43.2 ± 19.68	88.3 ± 7.53 ^c	42.0 ± 16.72

注：^b*P* < 0.01 vs 正常对照组；^c*P* < 0.05，^d*P* < 0.01 vs 模型对照组。

表 8-2 各组外周血 CD25⁺ FOXP3⁻/CD25⁻ FOXP3⁺/CD25⁻ FOXP3⁻ T 细胞水平的比较 (*n*=10, %, $\bar{x} \pm s$)

组 别	CD4 ⁺ CD25 ⁺ T 细胞	CD4 ⁺ FOXP3 ⁺ T 细胞	CD25 ⁺ FOXP3 ⁺ T 细胞
正常对照组	2.6 ± 0.97	30.5 ± 10.94	1.8 ± 0.27
模型组	3.8 ± 0.61 ^a	46.1 ± 3.03 ^b	1.2 ± 0.64 ^a
CⅡ组	2.4 ± 1.04 ^c	6.0 ± 1.31 ^d	1.6 ± 0.25 ^c
TWP 组	3.9 ± 1.64	50.9 ± 16.91	0.4 ± 0.57 ^c

注：^a*P* < 0.05，^b*P* < 0.01 vs 正常对照组；^c*P* < 0.05，^d*P* < 0.01 vs 模型对照组。

【案例辨析】 此试验涉及两个因素，即“是否 CIA 模型”和“灌胃溶液种类”。“是否 CIA 模型”为两水平试验因素；“灌胃溶液种类”是三水平处理因素，分别是“生理盐水”、“Ⅱ型胶原蛋白”、“雷公藤多苷”。而本研究只安排了四组：正常对照组、CIA 模型组、Ⅱ型胶原蛋白治疗组和雷公藤多苷对照组。本试验中没有安排对照组 + Ⅱ型胶原蛋白、对照组 + 雷公藤多苷。因而在实际分析中就不能将两因素之间可能存在的交互作用的效应反映出来。

【案例释疑】

(1)应该再安排两组：对照组 + II 型胶原蛋白、对照组 + 雷公藤多苷。这样就有 6 个组，为两个因素各有 2 个和 3 个水平的 6 种组合，这样的设计类型为两因素析因设计，不仅可以分析出两因素单独的效应，而且可以分析出因素之间可能存在的交互作用的效应大小。分析这种定量资料应该采用析因设计定量资料的方差分析，而不是用 *t* 检验。因此原表应该改成下面的格式，见表 8-3 和表 8-4。

表 8-3 各组外周血 CD4⁺ CD25⁺ FOXP3⁺ Treg 水平的比较(%, $\bar{x} \pm s$)

是否 CIA 模型	灌胃溶液种类	鼠 数	CD4 ⁺ CD25 ⁺ T 细胞	CD4 ⁺ FOXP3 ⁺ T 细胞	CD25 ⁺ FOXP3 ⁺ T 细胞
否	生理盐水	10	70.1 ± 12.52	93.2 ± 4.13	65.5 ± 12.64
	II 型胶原蛋白	×	×	×	×
	雷公藤多苷	×	×	×	×
是	生理盐水	10	59.9 ± 13.11	95.4 ± 2.21	49.3 ± 4.20
	II 型胶原蛋白	10	86.7 ± 15.32	97.1 ± 1.44	84.3 ± 15.81
	雷公藤多苷	10	43.2 ± 19.68	88.3 ± 7.53	42.0 ± 16.72

表 8-4 各组外周血 CD25⁺ FOXP3⁻ / CD25⁻ FOXP3⁺ / CD25⁻ FOXP3⁻ T 细胞水平的比较(%, $\bar{x} \pm s$)

是否 CIA 模型	灌胃溶液种类	鼠 数	CD25 ⁺ FOXP3 ⁻ T 细胞	CD25 ⁻ FOXP3 ⁺ T 细胞	CD25 ⁻ FOXP3 ⁻ T 细胞
否	生理盐水	10	2.6 ± 0.97	30.5 ± 10.94	1.8 ± 0.27
	II 型胶原蛋白	×	×	×	×
	雷公藤多苷	×	×	×	×
是	生理盐水	10	3.8 ± 0.61	46.1 ± 3.03	1.2 ± 0.64
	II 型胶原蛋白	10	2.4 ± 1.04	6.0 ± 1.31	1.6 ± 0.25
	雷公藤多苷	10	3.9 ± 1.64	50.9 ± 16.91	0.4 ± 0.57

(2)如果按原文的设计方案它属于多因素非平衡的组合试验，分析时对四组定量资料进行拆分，再选用合适的统计分析方法处理。可以拆分成以下 2 种组合：

组合 1：正常对照组和 CIA 模型组，见表 8-5 和表 8-6；

组合 2：CIA 模型组、II 型胶原蛋白治疗组和雷公藤多苷对照组，见表 8-7 和表 8-8。

表 8-5 正常对照组和 CIA 模型组各组外周血 CD4⁺ CD25⁺ FOXP3⁺ Treg 水平的比较($\bar{x} \pm s$)

分 组	鼠 数	CD4 ⁺ CD25 ⁺ T 细胞	CD4 ⁺ FOXP3 ⁺ T 细胞	CD25 ⁺ FOXP3 ⁺ T 细胞
正常对照组	10	70.1 ± 12.52	93.2 ± 4.13	65.5 ± 12.64
CIA 模型组	10	59.9 ± 13.11	95.4 ± 2.21	49.3 ± 4.20

表 8-6 正常对照组和 CIA 模型组各组外周血 CD25⁺ FOXP3⁻ / CD25⁻ FOXP3⁺ / CD25⁻ FOXP3⁻ T 细胞水平的比较($\bar{x} \pm s$)

分 组	鼠 数	CD25 ⁺ FOXP3 ⁻ T 细胞	CD25 ⁻ FOXP3 ⁺ T 细胞	CD25 ⁻ FOXP3 ⁻ T 细胞
正常对照组	10	2.6 ± 0.97	30.5 ± 10.94	1.8 ± 0.27
CIA 模型组	10	3.8 ± 0.61	46.1 ± 3.03	1.2 ± 0.64

表 8-7 正常对照组和 CIA 模型组各组外周血 CD4⁺ CD25⁺ FOXP3⁺ Treg 水平的比较($\bar{x} \pm s$)

分 组	鼠 数	CD4 ⁺ CD25 ⁺ T 细胞	CD4 ⁺ FOXP3 ⁺ T 细胞	CD25 ⁺ FOXP3 ⁺ T 细胞
CIA 模型组	10	59.9 ± 13.11	95.4 ± 2.21	49.3 ± 4.20
II 型胶原蛋白	10	86.7 ± 15.32	97.1 ± 1.44	84.3 ± 15.81
雷公藤多苷	10	43.2 ± 19.68	88.3 ± 7.53	42.0 ± 16.72

表 8-8 正常对照组和 CIA 模型组各组外周血 CD25⁺ FOXP3⁻/CD25⁻ FOXP3⁺/CD25⁻ FOXP3⁻ T 细胞水平的比较 ($\bar{x} \pm s$)

分 组	鼠 数	CD25 ⁺ FOXP3 ⁻ T 细胞	CD25 ⁻ FOXP3 ⁺ T 细胞	CD25 ⁻ FOXP3 ⁻ T 细胞
CIA 模型组	10	3.8 ± 0.61	46.1 ± 3.03	1.2 ± 0.64
Ⅱ型胶原蛋白	10	2.4 ± 1.04	6.0 ± 1.31	1.6 ± 0.25
雷公藤多苷	10	3.9 ± 1.64	50.9 ± 16.91	0.4 ± 0.57

第一种组合可以采用成组设计定量资料的 t 检验或单因素方差分析或 T^2 检验(多元方差分析);第二种组合可以采用单因素三水平设计定量资料一元或多元方差分析。注意定量资料应满足三个特定前提条件(即独立性、正态性和方差齐性),否则应该采用相应设计的非参数检验方法。

8.1.2 随机原则方面存在的问题

【案例 8-2】 原文目的观察试验性自身免疫性脑脊髓炎(EAE)发病高峰期淋巴细胞表面 B7-H1 分子的表达变化,探讨 B7-H1 在 EAE 发病和病情进展中的作用。用 6 周龄的 SPF 级野生型 C57BL/6 小鼠,雌性,体质量 20~25g,10 只。小鼠分组:EAE 组 5 只;未处理组 5 只。每天观察小鼠行为及神经功能表现,记录神经功能评分。

【案例辨析】 研究者对小鼠进行分组时,没有采用随机化方法。

【案例释疑】 应该采用随机化原则进行分组,使两对比组在重要的非试验因素方面具有极好的均衡性,提高可比性。

8.2 定量资料统计分析中存在的问题

8.2.1 误用成组设计定量资料 t 检验处理配对设计定量资料

【案例 8-3】 原文目的观察辅助性 17T(Th17)及其分泌的细胞因子 IL-17 在自身免疫性溶血性贫血(AIHA)患者中的变化并探讨其临床意义。研究对象选取 AIHA 患者 16 例作为病例组,对照组为 15 例正常人。检测对照组以及病例组经糖皮质激素治疗前后外周血中 Th17 细胞比率和血浆中 IL-17 的水平。统计方法两组间比较采用 t 检验,多组间比较采用方差分析, $P < 0.05$ 为差异有统计学意义,见表 8-9 和表 8-10。

表 8-9 各组外周血中 Th17 细胞的比率

组 别	n	Th17 细胞 (%)
病例组(治疗前)	16	2.78 ± 0.59 ^b
病例组(治疗后)	16	1.05 ± 0.28 ^{bd}
健康对照组	15	0.59 ± 0.15

注: ^b $P < 0.01$ vs 健康对照组; ^d $P < 0.01$ vs 本组治疗前。

表 8-10 各组血浆中 IL-17 水平的比较 ($\bar{x} \pm s$, ng/L)

组 别	n	IL-17 (ng/L)
病例组(治疗前)	16	126.4 ± 11.6 ^b
病例组(治疗后)	16	78.5 ± 6.4 ^{bd}
健康对照组	15	52.3 ± 4.8

注: ^b $P < 0.01$ vs 健康对照组; ^d $P < 0.01$ vs 本组治疗前。

【案例辨析】 原作者目的是比较对照组以及病例组经糖皮质激素治疗前后外周血中 Th17 细胞比率和血浆中 IL-17 的水平差异有无统计学意义,原作者主要用的是 t 检验。实际上根据分析的目的不同,将会提取不同组的资料进行比较,故各次分析所对应的设计类型是不一样的,所以采用的方法也不一样。例如,病例组治疗前或治疗后分别与健康对照组比较属于成组设计,可以考虑采用两独立样本 t 检验或两独立样本秩和检验,而病例组治疗前后的比较属于自身配对设计,应该采用配对 t 检验或配对符号秩检验。

【案例释疑】 病例组治疗前或治疗后分别与健康对照组比较时,先考察资料是否满足参数检验的

前提条件(即独立性、正态性和方差齐性),满足则可以采用成组设计的两独立样本 t 检验或单因素两水平设计定量资料方差分析,不满足则可做适当的变量变换或直接采用两独立样本比较的 Wilcoxon 秩和检验。病例组治疗前后的比较首先考察每对数据的差值是否满足正态性,满足则可采用配对设计定量资料 t 检验处理,如不满足正态性可采用配对设计定量资料 Wilcoxon 符号秩检验来处理。

值得注意的是,若同一个设计类型下有多个定量指标,当它们在专业上无联系时,可分别对每一个定量指标进行相应设计一元定量资料参数或非参数检验;当它们在专业上有联系且满足多元方差分析前提条件(方差和协方差矩阵满足正态性和方差齐性)时,可对它们同时进行相应设计定量资料多元方差或协方差分析,下同。

8.2.2 误用 t 检验处理单因素多水平设计定量资料

【案例 8-4】 原文目的探讨特发性血小板减少性紫癜(ITP)患者血清中血小板生成调控因子血小板生成素(TPO)和转化生长因子 $\beta 1$ (TGF- $\beta 1$) 的变化并探讨其临床意义。研究对象选取 30 例 ITP 患者,24 例病例对照和 25 例正常对照组。用 t 检验进行两组间差异比较, $P < 0.05$ 为有统计学意义,见表 8-11。

表 8-11 各组 TPO、TGF- $\beta 1$ 、PLT 监测结果 $\bar{x} \pm s$

组 别	n	TPO (ng/L)	TGF- $\beta 1$ (ng/L)	PLT
ITP 组	30	42.49 \pm 62.46	714.52 \pm 89.75 ^a	23.57 \pm 10.37
病例对照组	24	130.84 \pm 93.29 ^a	394.66 \pm 60.42	24.37 \pm 32.93
正常对照组	25	31.07 \pm 34.82	400.13 \pm 65.89	170.00 \pm 23.47

注: ^a $P < 0.05$ vs 正常对照组。

【案例辨析】 原作者目的是比较 ITP 患者、病例对照组和正常对照组三组间血清中血小板生成调控因子血小板生成素(TPO)和转化生长因子 $\beta 1$ (TGF- $\beta 1$) 水平的差异是否有统计学意义。原文设计类型为“单因素三水平设计”,而 t 检验只能用于不多于两组的数据比较。

【案例释疑】 在对原文三个组别间进行统计分析时,先考察该定量资料是否满足参数检验的前提条件(即独立性、正态性和方差齐性),如果满足则应采用单因素三水平设计定量资料方差分析,否则,可选用单因素三水平设计定量资料 Kruskal-Wallis H 检验。当三组间差别具有统计学意义时,才有必要进行两两比较。两两比较时不宜直接采用成组设计定量资料 t 检验,因为会增大犯 I 类错误的概率,应采用两两比较的方法,如 q 检验(SAS 软件中称为 SNK 法检验,属于参数检验)、Nemenyi 法检验(非参数检验)等。

【案例 8-5】 原文目的探讨瘦素(LP)在冠状动脉病变过程中的作用及临床价值,将 95 例冠心病患者根据冠状动脉造影检查结果分为 AMI 组、UAP 组和 SAP 组。检测血浆瘦素水平、C 反应蛋白(CRP)和白细胞介素-6(IL-6)。统计方法计量资料比较采用 t 检验, $P < 0.05$ 为差异具有统计学意义。冠心病各组患者瘦素和炎性水平结果比较显示:UAP 组瘦素和炎性水平显著高于 SAP 组,有显著性差异;AMI 组瘦素和炎性水平显著高于 UAP 组,相比较差异有统计学差异,见表 8-12。

表 8-12 冠心病各组患者瘦素和炎性水平结果比较 ($\bar{x} \pm s$)

组 别	n	LP ($\mu\text{g/L}$)	CRP (mg/L)	IL-6 ($\mu\text{g/L}$)
SAP 组	30	7.12 \pm 2.78	9.64 \pm 4.23	39, 41 \pm 12.81
UAP 组	31	9.03 \pm 3.14 ^a	11.15 \pm 5.57 ^a	50.18 \pm 14.45 ^a
AMI 组	34	12.87 \pm 5.19 ^a	14.67 \pm 6.08 ^a	67.45 \pm 17.72 ^a

注: ^a $P < 0.05$ 。

【案例辨析】 原文目的是比较 AMI 组、UAP 组和 SAP 组三组间瘦素和炎性水平差异是否有统计学意义,原作者采用 t 检验是错误的, t 检验只能用于不多于两组的数据比较。

【案例释疑】 原文设计类型为“单因素三水平设计”,三组间进行统计分析时,先考察该资料是否满足参数检验的前提条件,如果满足则应采用单因素三水平设计定量资料方差分析,否则,可选用单因素三水平设计定量资料非参数检验。当三组间差别具有统计学意义时,再进行两两组间比较,而不宜直接采用 t 检验,因为那样做会增大犯 I 类错误的概率。

8.2.3 误用单因素多水平设计定量资料方差分析处理多因素析因设计定量资料

【案例 8-6】 原文目的观察热休克蛋白 27 (Hsp27) 对大鼠臂丛神经根撕脱后脊髓运动神经元一氧化氮合酶(NOS)表达的影响。方法:将 Wistar 大鼠 60 只,随机分为试验组和对照组,两组分别于术后 12h、1d、3d、5d、7d 5 个时间点处死动物(每个时间点 6 只)。进行脊髓前角阳性细胞计数和染色强度测定。统计分析采用单因素方差分析, $P < 0.05$ 表示有统计学意义,见表 8-13 和表 8-14。

表 8-13 脊髓前角 Hsp27 阳性神经元数目及平均吸光度值 (A) ($n=6, \bar{x} \pm s$)

时 间	阳性神经元数	A 值	阳性神经元数	A 值
	试 验 组		对 照 组	
12h	16.34 \pm 2.36 ^b	0.130 \pm 0.012 ^b	8.27 \pm 1.81	0.082 \pm 0.011
1d	20.75 \pm 4.59 ^b	0.212 \pm 0.034 ^b	12.31 \pm 2.06	0.157 \pm 0.024
3d	17.51 \pm 2.39 ^b	0.126 \pm 0.021 ^a	10.65 \pm 1.28	0.106 \pm 0.013
5d	15.47 \pm 4.10 ^b	0.118 \pm 0.014 ^a	8.01 \pm 1.47	0.089 \pm 0.010
7d	12.19 \pm 2.91 ^a	0.102 \pm 0.012	7.69 \pm 1.04	0.075 \pm 0.012

注:^a $P < 0.05$,^b $P < 0.05$ vs 对照组。

表 8-14 脊髓前角 NADPH-d 阳性神经元数目及平均 A 值 ($n=6, \bar{x} \pm s$)

时 间	阳性神经元数	A 值	阳性神经元数	A 值
	试 验 组		对 照 组	
12h	14.34 \pm 5.26	0.193 \pm 0.017	—	0.109 \pm 0.016
1d	9.57 \pm 3.12	0.129 \pm 0.032	7.25 \pm 3.08	0.125 \pm 0.205
3d	9.96 \pm 2.01	0.126 \pm 0.020	9.36 \pm 2.61	0.128 \pm 0.017
5d	9.39 \pm 2.73	0.132 \pm 0.017 ^a	11.48 \pm 2.09	0.156 \pm 0.036
7d	9.32 \pm 2.61	0.138 \pm 0.023 ^a	12.79 \pm 1.64	0.172 \pm 0.013

注:^a $P < 0.05$ vs 对照组。

【案例辨析】 从原文看出本研究试验设计有两个因素,一个试验因素是组别,有 2 个水平(试验组和对照组);另外一个试验因素是时间,有 5 个水平(术后 12h、1d、3d、5d、7d),而且这里的时间因素不是重复测量因素,因为不同时间点用的大鼠是不同的,所以不同时间点的数据是独立的。本资料是个涉及两因素的试验,因此原作者采用单因素方差分析是错误的。

【案例释疑】 该资料设计类型为两因素析因设计,若资料满足参数检验的前提条件,则采用两因素析因设计定量资料方差分析来处理,可以比较试验组和对照组间差异是否有统计意义,不同时间点间差异是否有统计意义,还可以分析两者是否存在交互作用。若不满足参数检验的前提条件,应进行适当的变量变换,当变换后的资料满足参数检验的前提条件时,对变换后的资料采用两因素析因设计定量资料方差分析来进行处理。

【案例 8-7】 原文目的观察信号转导子与转录活化子 1 (STAT1) 反义寡核苷酸 (ASON) 雾化吸

入对博莱霉素 (BLM) 致肺纤维化大鼠肺泡灌洗液 (BALF) PDGF-BB (血小板衍生生长因子 - BB)、TNF- α (肿瘤坏死因子 - α)、TGF- β 1 (转化生长因子 - β 1)、IFN- γ (干扰素 - γ) 表达及肺组织 I、III 型胶原 mRNA 表达的影响。方法: 取 45 只健康雌性 Wistar 大鼠随机分成 ASON 组、BLM 组和生理盐水 (NS) 组各 15 只, 气管内分别灌注 BLM (ASON 组、BLM 组) 和 NS (NS 组) 复制肺纤维化模型和空白对照, ASON 组于第 0、2、4、6 天雾化吸入 STAT1 ASON, BLM 组和 NS 组雾化吸入 NS, 各组分别于第 7、14、28 天均处死动物 5 只。收集 BALF 测定 PDGF-BB、TNF- α 、IFN- γ 、TGF- β 1 的水平。样本均数的比较采用单因素方差分析和 q 检验, $P < 0.05$ 为差异有统计学意义。具体见表 8-15、表 8-16、表 8-17 和表 8-18。

表 8-15 STAT1 反义寡核苷酸对肺纤维化大鼠 BALF 中 TGF- β 1 含量的影响 ($n=5$, ng/L, $\bar{x} \pm s$)

Group	7d	14d	28d
NS	42.53 \pm 1.88	43.08 \pm 3.20	41.57 \pm 2.89
BLM	66.43 \pm 7.97 ^a	63.50 \pm 7.91 ^a	57.67 \pm 2.46 ^a
ASON	55.37 \pm 4.46 ^{a,c}	50.30 \pm 2.57 ^{a,c}	49.51 \pm 3.47 ^{a,c}

注: ^a $P < 0.05$ vs NS group; ^c $P < 0.05$ vs BLM group。

表 8-16 STAT1 反义寡核苷酸对肺纤维化大鼠 BALF 中 PDGF-BB 含量的影响 ($n=5$, ng/L, $\bar{x} \pm s$)

Group	7d	14d	28d
NS	25.61 \pm 4.02	26.54 \pm 6.22	25.72 \pm 4.58
BLM	111.38 \pm 8.19 ^a	90.50 \pm 6.61 ^a	56.26 \pm 6.85 ^a
ASON	65.17 \pm 7.26 ^{a,c}	50.86 \pm 5.49 ^{a,c}	43.12 \pm 4.66 ^{a,c}

注: ^a $P < 0.05$ vs NS group; ^c $P < 0.05$ vs BLM group。

表 8-17 STAT1 反义寡核苷酸对肺纤维化大鼠 BALF 中 TNF- α 含量的影响 ($n=5$, ng/L, $\bar{x} \pm s$)

Group	7d	14d	28d
NS	1.610 \pm 0.31	1.44 \pm 0.26	1.48 \pm 0.40
BLM	2.985 \pm 0.40 ^a	2.85 \pm 0.24 ^a	2.45 \pm 0.25 ^a
ASON	2.250 \pm 0.36 ^{a,c}	2.05 \pm 0.22 ^{a,c}	1.93 \pm 0.14 ^{a,c}

注: ^a $P < 0.05$ vs NS group; ^c $P < 0.05$ vs BLM group。

表 8-18 STAT1 反义寡核苷酸对肺纤维化大鼠 BALF 中 IFN- γ 含量的影响 ($n=5$, ng/L, $\bar{x} \pm s$)

Group	7d	14d	28d
NS	73.03 \pm 2.35	75.83 \pm 2.44	76.86 \pm 4.58
BLM	65.62 \pm 3.88 ^a	48.35 \pm 3.06 ^a	43.30 \pm 3.66 ^a
ASON	73.55 \pm 2.67 ^c	63.32 \pm 1.97 ^{a,c}	54.87 \pm 2.92 ^{a,c}

注: ^a $P < 0.05$ vs NS group; ^c $P < 0.05$ vs BLM group。

【案例辨析】 从原文看出本研究试验设计有两个因素, 一个试验因素是组别, 有 3 个水平 (NS 组、BLM 组和 ASON 组); 另外一个试验因素是时间, 也有 3 个水平 (7d、14d、28d), 而且这里的时间因素不是重复测量因素, 因为不同时间点用的大鼠是不同的, 所以不同时间点的数据是独立的。本资料是个涉及两因素的试验, 因此原作者采用单因素方差分析是错误的。

【案例释疑】 该资料设计类型为两因素析因设计, 若资料满足参数检验的前提条件, 则采用两因素析因设计定量资料方差分析来处理, 可以比较组别不同水平之间差异是否有统计意义, 不同时间点间差异是否有统计意义, 还可以分析二者间的交互作用。当交互作用有统计学意义时, 还可以将其中一个因素依次控制在不同水平上, 对另一个因素不同水平间进行两两比较。若不满足

参数检验的前提条件,应进行适当的变量变换,当变换后的资料满足参数检验的前提条件时,对变换后的资料进行两因素析因设计定量资料方差分析。

【案例8-8】 原文目的观察缺血再灌注损伤(IRI)后,低龄和高龄 p53(+/+)和 p53(-/-)鼠肾小管上皮细胞的演变,探讨 p53 基因对 IRI 后肾小管上皮细胞演变的影响。方法:低龄 p53(+/+)和 p53(-/-)鼠、高龄 p53(+/+)和 p53(-/-)鼠四组动物(每组 35 只),分别在 IRI 后 0 天、1 天、3 天、7 天、1 月、3 月和 6 月各处死 5 只收集肾脏标本。HE 染色后光镜下观察 IRI 后肾小管坏死的变化。统计方法多组资料间和同组多个时点资料间比较,方差齐者采用 One-Way ANOVA 分析,组间两两比较采用 LSD 法,方差不齐者采用 Tamhane 分析。 $P < 0.05$ 为具有统计学意义。结果显示:IRI 后 0d,肾小管以坏死为主,高龄鼠比低龄鼠明显($P < 0.05$),p53(-/-)鼠比 p53(+/+)鼠明显($P < 0.05$)。具体见表 8-19。

表 8-19 缺血再灌注损伤后肾小管坏死的评估($n=5$, $\bar{x} \pm s$)

Time After IRI	p53 (+ / +) mice 的 $\bar{x} \pm s$		p53 (- / -) mice 的 $\bar{x} \pm s$	
	Young mice	Old mice	Yong mice	Old mice
0d	1.7 \pm 0.7	2.4 \pm 0.6 ^a	2.3 \pm 0.1 ^c	3.1 \pm 0.7 ^{ac}
1d	1.8 \pm 0.8	2.1 \pm 0.8 ^a	2.2 \pm 0.5 ^c	2.7 \pm 0.4 ^{ac}
3d	1.2 \pm 0.5	1.8 \pm 0.8 ^a	1.8 \pm 0.5 ^c	2.3 \pm 0.3 ^a
7d	0.8 \pm 0.7	1.0 \pm 0.7 ^a	1.4 \pm 0.1 ^c	1.9 \pm 0.1 ^{ac}
1 month	0.2 \pm 0.1	0.3 \pm 0.1	0.4 \pm 0.2	0.5 \pm 0.2
3 month	0.0 \pm 0.0	0.0 \pm 0.0	0.0 \pm 0.0	0.2 \pm 0.0
6 month	ND	ND	0.0 \pm 0.0	0.2 \pm 0.0

注: ^a $P < 0.05$ vs young mice with the same genotype and at the same time; ^c $P < 0.05$ vs P53 (+ / +) mice at the same age and the same time; ND: No detection。

【案例辨析】 从原文看出本研究试验设计有三个因素,一个试验因素是月龄,有 2 个水平(低龄组和高龄组),另外一个试验因素是 p53,也有 2 个水平[p53(+/+)组和 p53(-/-)组],最后一个试验因素是时间,有 7 个水平(IRI 后 0 天、1 天、3 天、7 天、1 月、3 月和 6 月),而且这里的时间因素不是重复测量因素,因为不同时间点用的鼠是不同的,所以不同时间点的数据是独立的。本资料是个涉及三因素的试验,因此原作者采用单因素方差分析是错误的。

【案例释疑】 该资料设计类型为三因素析因设计,若资料满足参数检验的前提条件,则采用三因素析因设计定量资料方差分析来处理。可以分析 p53(+/+)鼠和 p53(-/-)鼠之间的差异,低龄鼠和高龄鼠之间的差异,不同时间点的差异,还可以分析它们之间的交互作用。若不满足参数检验的前提条件,应进行适当的变量变换,当变换后的资料满足参数检验的前提条件时,对变换后的资料采用三因素析因设计定量资料方差分析来进行处理。

值得注意的是,以上三个例子都被判定为析因设计,其隐含的两个前提条件是全部试验因素在进行试验时同时施加(若试验因素施加有先后顺序,则应判定为裂区或分割设计)且试验因素对观测结果的影响地位平等(若试验因素存在自然属性上的嵌套关系或有主次之分,则应判定为嵌套或系统分组设计)。

8.2.4 误用 t 检验处理具有一个重复测量单因素设计定量资料

【案例8-9】 原文目的探讨围生期孕妇血清 IL-6、TNF- α 及外周血 T 淋巴细胞亚群变化规律。选取 100 例产妇为研究对象,将其分别于分娩发动前、妊娠发动时及分娩后 1d、3d 的血清 IL-6、TNF- α 及外周血 T 淋巴细胞亚群水平进行比较。统计方法采用 t 检验,以 $P < 0.05$ 为有统计学意义具体情况见表 8-20 和表 8-21。

表 8-20 围生期孕妇血清 IL-6、TNF-α 水平变化比较

时 间	IL-6 (ng/L)	TNF-α (μg/L)	时 间	IL-6 (ng/L)	TNF-α (μg/L)
分娩前	148.67 ± 22.98	0.48 ± 0.17	妊娠发动时	1058.42 ± 82.36	0.86 ± 0.15
分娩后 1d	521.47 ± 58.74	0.69 ± 0.20	分娩后 3d	156.93 ± 33.86	0.42 ± 0.19

表 8-21 围生期孕妇外周血 T 淋巴细胞亚群变化比较 (%)

时 间	CD3 +	CD4 +	CD8 +	CD4/CD8
分娩前	78.78 ± 2.56	37.25 ± 4.31	25.15 ± 4.06	1.69 ± 0.33
妊娠发动时	68.73 ± 3.12	28.57 ± 5.13	21.52 ± 3.43	1.42 ± 0.28
分娩后 1d	71.58 ± 2.67	33.56 ± 4.87	23.30 ± 3.19	1.55 ± 0.36
分娩后 3d	77.54 ± 3.52	36.96 ± 5.45	25.20 ± 2.98	1.60 ± 0.29

【案例辨析】 原文目的是比较分娩发动前、妊娠发动时及分娩后 1d、3d 的血清 IL-6、TNF-α 及外周血 T 淋巴细胞亚群水平。该研究设计属于重复测量的单因素设计，时间是重复测量因素，有 4 个水平(分娩发动前、妊娠发动时及分娩后 1d、3d)，采用 *t* 检验分析重复测量设计定量资料是错误的。

【案例释疑】 该资料试验设计类型为“具有一个重复测量的单因素设计”，可以近似采用随机区组设计定量资料方差分析处理该重复测量设计定量资料，但它要求资料的协方差矩阵必需满足 H 型条件($\sigma_{jj} + \sigma_{kk} - 2\sigma_{jk} = c$ ，对所有 *j*、*k* 成立，其中 *c* 为常数)，当资料的协方差矩阵不满足 H 型条件时，需计算校正系数，用它来对相应的自由度作校正。可用 Mauchly 球性检验的方法来检验协方差矩阵 H 型条件。不同时间点的两两比较：如果协方差矩阵基本符合球性条件时用 Tukey 法，如果偏离球性条件较远时用 Bonferroni 法。

8.2.5 误用 *t* 检验处理具有一个重复测量两因素设计定量资料

【案例 8-10】 原文目的探讨肿瘤标志物细胞角蛋白片段抗原 21-1 (CYFRA 21-1)、人体表皮生长因子受体-2 (HER-2/neu) 和神经烯醇化酶 (NSE) 对良、恶性胸水鉴别诊断的临床应用价值。方法：定量检测 62 例恶性及 48 例良性胸腔积液患者血清和胸腔积液中 CYFRA 21-1、HER-2/neu 和 NSE 的含量。组间比较采用 *t* 检验，率的比较采用 χ^2 检验，*P* < 0.05 表示差异有统计学意义。结果显示：恶性组胸水及血清 CYFRA 21-1、HER-2/neu 和 NSE 的含量均明显高于良性组 (*P* < 0.01)。恶性组患者胸水中 3 种肿瘤标志物的含量均显著高于血清中的含量 (*P* < 0.05 或 0.01)，见表 8-22。

表 8-22 两组患者血清及胸水中 3 种肿瘤标志物含量的检测 ($\bar{x} \pm s$, μg/L)

组 别	NSE		CYFRA 21-1		HER-2/neu	
	血 清	胸 水	血 清	胸 水	血 清	胸 水
恶性组 (<i>n</i> = 62)	18.14 ± 16.83 ^b	25.61 ± 19.06 ^{a,b}	12.64 ± 9.63 ^b	19.06 ± 14.92 ^{b,d}	3.85 ± 1.28 ^b	7.18 ± 4.21 ^{b,d}
良性组 (<i>n</i> = 48)	7.90 ± 2.47	8.75 ± 3.87	2.93 ± 1.30	3.42 ± 1.70	1.85 ± 1.17	2.14 ± 0.68

注：^a*P* < 0.05，^d*P* < 0.01 vs 血清组；^b*P* < 0.01 vs 良性组。

【案例辨析】 原文进行了恶性组和良性组间 3 种肿瘤标志物的比较，同时也进行了胸水和血清间 3 种肿瘤标志物的比较。由此可知本研究是一个涉及两因素的试验，一个因素是组别，有两个水平(恶性组和良性组)，另外一个因素是部位，也有两个水平(血清和胸水)。而且部位因素是个重复测量因素，因为血清和胸水的数据来自同一人，它们之间是不独立的。因此本资料不能用 *t* 检验进行两组间的比较，那样做破坏了整体设计。

【案例释疑】该资料设计类型为“重复测量的两因素设计”，若资料满足参数检验的前提条件，可采用该设计定量资料方差分析来处理，可以分析恶性组和良性组之间的差异，胸水和血清之间的差异，以及这两个试验因素间的交互作用。

【案例8-11】原文目的探讨苦参素联合甘利欣治疗慢性乙型肝炎的疗效，及对肝纤维化指标的影响。方法：选取80例慢性乙型肝炎合并肝纤维化患者作为研究对象，随机分成治疗组和对照组各40例，治疗组给予甘利欣加苦参素，对照组应用复方丹参治疗。观察比较两组的临床疗效及肝功能指标与肝纤维化指标的改变。结果显示治疗前两组的肝纤维化指标HA、LN、PCⅢ、IV-C比较，差异无显著性，治疗3个月后，肝纤维化指标HA、LN、PCⅢ、IV-C均较治疗前明显下降，且治疗组与对照组比较，差异有统计学意义($P < 0.05$)。治疗前两组的肝功能指标ALT、AST、TBIL、ALB比较，差异无显著性，治疗3个月后，肝功能指标ALT、AST、TBIL、ALB比较均较治疗前明显下降，但治疗组与对照组比较，差异无显著性($p > 0.05$)。统计方法率的比较采用卡方检验， $P < 0.05$ 具有统计学意义。具体比较见表8-23和表8-24。

表8-23 两组治疗前后肝功能指标比较($n=40$, $\bar{x} \pm s$)

组别		ALT (U/L)	AST (U/L)	TBIL ($\mu\text{mol/L}$)	ALB (g/L)
治疗组	治疗前	296.6 \pm 165.4	114 \pm 55.7	65.4 \pm 21.7	32.2 \pm 2.7
	治疗后	57.7 \pm 26.3 ^b	36 \pm 17.5 ^b	20.3 \pm 9.5 ^b	37.6 \pm 3.6 ^b
对照组	治疗前	288.4 \pm 173.6	116 \pm 58.2	63.4 \pm 23.3	34.4 \pm 2.5
	治疗后	168.3 \pm 55.8	74 \pm 25.5	54.5 \pm 24.4	32.0 \pm 5.3

注：^b $P < 0.01$ vs 治疗组的治疗前。

表8-24 两组治疗前后肝纤维化指标比较($n=40$, $\mu\text{g/L}$, $\bar{x} \pm s$)

组别		HA	LN	PCⅢ	IV-C
治疗组 ^a	治疗前	336.7 \pm 155.7	165.6 \pm 85.6	190.3 \pm 111.6	132.4 \pm 102.4
	治疗后	94.3 \pm 22.3	74.6 \pm 14.4	95.6 \pm 32.5	96.6 \pm 34.6
对照组	治疗前	324.0 \pm 158.2	165.7 \pm 87.2	192.3 \pm 110.3	131.4 \pm 101.7
	治疗后	187.3 \pm 97.4	121.4 \pm 31.0	142.0 \pm 59.6	113.5 \pm 45.5

注：^a $P < 0.01$ vs 对照组。

【案例辨析】由原文可知本研究是个涉及两因素的试验，一个因素是组别，有两个水平(治疗组和对照组)，另外一个因素是时间，也有两个水平(治疗前和治疗后)，时间因素是个重复测量因素。原文中没有说明定量资料的统计分析方法。

【案例释疑】该资料设计类型为“具有一个重复测量的两因素设计”，可以考虑用此设计定量资料方差分析来处理，可以分析治疗组和对照组之间是否有差异、治疗前后是否有差异，还可以分析两者间的交互作用。

8.2.6 误用 t 检验和单因素多水平设计定量资料方差分析处理多因素重复测量设计定量资料

【案例8-12】原文目的探讨2000 $\mu\text{W}/\text{cm}^2$ 电离辐射对大鼠海马N-甲-D-门冬氨酸受体亚单位2A(NR2A)蛋白及其mRNA水平表达的影响，揭示电磁辐射对大鼠学习记忆功能的损伤机制。方法：试验分为空白对照组、假辐射组、1h/d、2h/d、3h/d辐射组，通过Morris水迷宫检测大鼠的空间学习记忆能力。统计处理组间差异显著性用方差分析法检验，组间两两比较用LSD- t 检验， $P < 0.05$ 为差异有统计学意义，见表8-25。

表 8-25 各组大鼠寻找安全平台的平均潜伏期 ($n=20, \bar{x} \pm s$)

组 别	平均潜伏期 (s)			
	第 1 天	第 2 天	第 3 天	第 4 天
空白对照组	76.28 ± 16.29	39.98 ± 7.93	19.12 ± 3.39	8.80 ± 1.66
假辐射组	75.13 ± 15.68	41.16 ± 8.13	22.86 ± 4.25	8.96 ± 1.58
1h/d 组	77.69 ± 16.39	53.68 ± 10.23 ^a	24.68 ± 4.46 ^a	12.29 ± 1.36 ^a
2h/d 组	77.27 ± 17.11	56.82 ± 11.22 ^a	28.18 ± 5.59 ^a	17.99 ± 2.25 ^b
3h/d 组	79.88 ± 19.39	62.28 ± 10.89 ^a	40.18 ± 7.79 ^b	24.66 ± 5.56 ^b

注: ^a $P < 0.05$, ^b $P < 0.01$ vs 对照组。

【案例辨析】 原作者目的是考察 $2000 \mu\text{W}/\text{cm}^2$ 电离辐射对大鼠学习记忆能力的影响,意在比较各组别大鼠寻找安全平台的平均潜伏期是否有差异。原文把训练时间固定在每一天,然后分别用其他四组和空白对照组作比较,用了单因素方差分析和两两比较中的 LSD- t 检验。从原文设计可知训练时间也应该看成一个试验因素来处理,而原文只是在训练时间的每个水平下进行组间的比较,因此是严重错误的。此试验涉及两个因素,一个因素是分组因素,有 5 个水平,分别是空白对照组、假辐射组、1h/d、2h/d、3h/d 辐射组;另外一个因素是训练时间,有 4 个水平,分别是第 1 天、第 2 天、第 3 天、第 4 天。而且从原文判断试验因素训练时间是个典型的重复测量因素,所以不能采用析因设计两因素方差分析,必须采用具有一个重复测量设计两因素定量资料方差分析来进行。

【案例释疑】 该资料试验设计类型为“具有一个重复测量两因素设计”,若资料满足参数检验的前提条件,则采用此设计定量资料方差分析。可以比较各组别之间大鼠寻找安全平台的平均潜伏期是否有差异,可以比较不同训练时间之间大鼠寻找安全平台的平均潜伏期是否有差异,还可以分析二者间是否存在交互作用,也可以比较任两种组合(组别和时间)之间的差异。若资料不满足参数检验的前提条件,应做适当的变量变换,当变换后的资料满足参数检验的前提条件后,对变换后的资料采用此设计定量资料方差分析处理。

值得注意的是,严格地说,本例中的“组别”并非是一个具有 5 水平的试验因素,而是某两个试验因素水平的非全面组合的结果,拆分“组别”后再判定设计类型更科学严谨。前两行为一个组合(分组因素可被称为“是否接受假照射”);后 4 行为一个组合(分组因素可被称为“每天照射的剂量”,其中假照射组可被视为“0h/d 组”)。与重复测量时间因素联合起来考虑,两种组合的设计名称相同,都是具有一个重复测量两因素设计。

【案例 8-13】 原文目的观察脑出血患者 CSF 和血清中 S100 β 蛋白含量的变化及其临床意义。将 24 只家兔随机分为两组(每组 12 只),一组为试验组,另一组为假手术组。分别于试验性脑出血后 6h、12h、24h、48h、72h、96h 等各时间点取各组 CSF 和血清,测定 S100 β 蛋白含量。统计分析用 SPSS 10.0 软件,计量资料两两间比较采用 t 检验。具体比较见表 8-26 和表 8-27。

表 8-26 不同时间点两组动物 CSF 中 S100 β 蛋白水平的比较

组 别	建模手术后不同时间 (h) 的 $\bar{x} \pm s$					
	6	12	24	48	72	96
试验组 ($n=11$)	53.2 ± 5.6 ^a	113.7 ± 5.5 ^b	204.9 ± 5.1 ^b	362.5 ± 6.0 ^b	451.6 ± 5.4 ^b	378.1 ± 5.9 ^b
假手术组 ($n=12$)	48.6 ± 2.1	52.3 ± 2.3	45.6 ± 2.2	47.2 ± 2.7	54.8 ± 2.5	51.3 ± 2.4

注: ^a $P < 0.05$, ^b $P < 0.01$ vs 假手术组。

【案例辨析】 原作者目的是比较不同时间点两组动物 CSF 及血清中 S100 β 蛋白检测水平。从试验设计类型来看有两个试验因素,一个试验因素是组别,它有两个水平(试验组和假手术组),另外一

个试验因素是建模手术后不同时间,它有6个水平(6h、12h、24h、48h、72h和96h),而且它是个重复测量因素。因此原文固定在每个时间点采用 t 检验比较试验组和假手术组的差异是错误的。

表 8-27 不同时间点两组动物血清中 S100 β 蛋白水平的比较

组 别	建模手术后不同时间 (h) 的 $\bar{x} \pm s$					
	6	12	24	48	72	96
试验组 ($n=11$)	20.2 ± 3.6^a	43.7 ± 5.5^b	98.9 ± 4.1^b	168.5 ± 5.1^b	356.6 ± 3.4^b	264.1 ± 4.7^b
假手术组 ($n=12$)	8.6 ± 1.1	12.3 ± 2.1	25.4 ± 1.2	35.2 ± 2.3	44.2 ± 2.3	37.3 ± 2.3

注: $^aP < 0.05$, $^bP < 0.01$ vs 假手术组。

【案例释疑】 该资料试验设计类型为“具有一个重复测量两因素设计”,若资料满足参数检验的前提条件,则采用此设计定量资料方差分析。可以比较试验组和假手术组之间 S100 β 蛋白含量是否有差异,可以比较不同时间点之间 S100 β 蛋白含量是否有差异,还可以分析二者间的交互作用。若资料不满足参数检验的前提条件,应做适当的变量变换,当变换后的资料满足参数检验的前提条件后,对变换后的资料进行此设计定量资料方差分析。

【案例 8-14】 原文目的探讨连续性血液净化(CBP)治疗对多发伤后并发脓毒症患者血清血管生成素(Ang-2)、TNF- α 与 IL-18 水平及预后的影响。选择 73 例多发伤后并发脓毒症患者,按照是否接受连续性血液净化治疗分为常规治疗组 35 例和连续性血液净化治疗组 38 例,监测治疗前后临床疗效及预后,并分别在治疗前和治疗后 6h、12h、24h、48h 各时间点检测两组患者血清 Ang-2、TNF- α 与 IL-18 水平的变化。结果显示经 72h 治疗后, CBP 治疗组患者血清 Ang-2、TNF- α 与 IL-18 水平均较治疗前明显降低($P < 0.05$),而常规治疗组患者在治疗 72h 内血清 Ang-2、TNF- α 、IL-18 水平均较治疗前无明显变化($P > 0.05$)。组间比较, CBP 治疗组患者的各时间点血清 Ang-2、TNF- α 、IL-18 水平均明显低于常规治疗组。统计分析组间比较采用 t 检验,计量资料采用 χ^2 检验。 $P < 0.05$ 表示差异有统计学意义,见表 8-28。

表 8-28 两组治疗前后血清细胞因子水平变化比较($\bar{x} \pm s$)

组 别	时 间 点	Ang-2 ($\mu\text{g/L}$)	TNF- α ($\mu\text{g/L}$)	IL-18 (ng/L)
常规治疗组 ($n=35$)	治疗前	12.46 ± 6.35	482.57 ± 142.79	592.67 ± 227.39
	治疗 6h	12.32 ± 5.78	472.39 ± 132.45	585.46 ± 205.49
	治疗 12h	12.19 ± 5.42	469.94 ± 121.36	582.52 ± 242.85
	治疗 24h	12.27 ± 5.96	465.57 ± 125.69	589.36 ± 182.59
	治疗 48h	12.35 ± 6.41	476.43 ± 134.87	598.47 ± 258.28
CBP 治疗组 ($n=38$)	治疗前	12.53 ± 6.48	469.26 ± 132.84	590.84 ± 218.26
	治疗 6h	8.79 ± 5.47^a	378.25 ± 112.46^a	476.75 ± 192.58^a
	治疗 12h	7.25 ± 4.82^a	286.53 ± 92.59^a	392.50 ± 152.48^a
	治疗 24h	6.69 ± 3.96^a	196.50 ± 62.57^a	315.72 ± 138.54^a
	治疗 48h	5.65 ± 3.41^a	125.93 ± 27.19^a	254.37 ± 142.59^a

注: $^aP < 0.05$ vs 治疗前和常规治疗组。

【案例辨析】 首先原文统计分析方法表述为计量资料采用 χ^2 检验,这是很明显的错误。原作者目的是比较常规治疗组和 CBP 治疗组在不同时间点(治疗前后)血清 Ang-2、TNF- α 、IL-18 水平是否有差异。该试验共涉及两个试验因素,一个试验因素是组别,有两个水平(常规治疗组和 CBP 治疗组);另一个试验因素是不同时间点(治疗前后),有 5 个水平(治疗前、治疗后 6h、12h、24h、48h),而且不同时间点是重复测量因素,因此原作者采用 t 检验分析该资料是错误的。

【案例释疑】 该资料试验设计类型为“具有一个重复测量两因素设计”,若资料满足参数检验

的前提条件,则采用此设计定量资料方差分析。可以比较常规治疗组和 CBP 治疗组之间血清 Ang-2、TNF- α 、IL-18 水平是否有差异,可以比较不同时间点(治疗前后)之间血清 Ang-2、TNF- α 、IL-18 水平是否有差异,还可以分析二者间的交互作用,也可以比较任两种组合(组别和时间点)之间的差异。若资料不满足参数检验的前提条件,应做适当的变量变换,当变换后的资料满足参数检验的前提条件后,对变换后的资料采用此设计定量资料方差分析处理。

【案例 8-15】 原文目的建立环磷酰胺(CTX)治愈荷膀胱癌 T739 小鼠的动物模型,观察外周血象特别是血小板计数在化疗治愈肿瘤中的动态变化及其意义。方法:取肿瘤接种后第 7 天的荷瘤小鼠 12 只,称重后随机分组,每组 6 只,最低有效量 CTX 单次荷瘤小鼠腹腔内注射,等量生理盐水对照。分别在用药后 6h 及第 2、4、9、14 天行内眦静脉取血,进行血常规检测,了解 CTX 治疗后外周血红细胞(RBC)、血红蛋白(Hb)、白细胞(WBC)和血小板(Plt)计数的动态变化。试验结果采用 t 检验。结果显示 15mg/kg CTX 单次腹腔内注射在治愈多数荷瘤小鼠的同时对外周血象无明显抑制作用。用药后 6h 治疗组荷瘤鼠外周血血小板计数明显高于对照组小鼠,两组数据比较具有统计学意义($P < 0.01$);用药后第 2、4、9、14 天治疗组荷瘤鼠外周血血小板计数与对照组小鼠比较差异无统计学意义($P > 0.05$),见表 8-29。

表 8-29 治疗后荷瘤小鼠外周血全血细胞的动态变化($n=6, \bar{x} \pm s$)

Group	Time after drug injection	Hb (g/L)	RBC ($\times 10^{12}/L$)	WBC ($\times 10^9/L$)	Plt ($\times 10^9/L$)
NS control	6h	148.00 \pm 10.0	9.00 \pm 0.60	10.54 \pm 1.74	1086.6 \pm 81.0
	2d	105.34 \pm 9.46	6.86 \pm 0.64	16.86 \pm 4.64	474.0 \pm 37.4
	4d	91.34 \pm 3.06	5.78 \pm 0.20	11.14 \pm 2.14	927.4 \pm 119.8
	9d	114.70 \pm 10.08	7.42 \pm 0.20	11.86 \pm 2.70	970.0 \pm 56.4
	14d	139.00 \pm 15.56	8.88 \pm 1.14	9.86 \pm 3.40	810.0 \pm 189.6
CTX treatment	6h	166.60 \pm 15.20	9.86 \pm 0.64	12.38 \pm 3.96	1483.4 \pm 184.4 ^a
	2d	114.66 \pm 3.06	7.36 \pm 0.18	13.66 \pm 1.64	532.6 \pm 48.4
	4d	99.34 \pm 4.16 ^a	6.14 \pm 0.20	11.36 \pm 0.30	841.4 \pm 124.4
	9d	105.34 \pm 11.00	7.06 \pm 0.68	12.50 \pm 4.28	887.4 \pm 279.0
	14d	108.00 \pm 14.42 ^a	7.68 \pm 0.56	8.46 \pm 0.76	513.0 \pm 250.4

注: ^a $P < 0.05$ vs NS control group。

【案例辨析】 从试验设计类型来看有两个试验因素,一个试验因素是组别,它有 2 个水平(对照组和 CTX 组),另外一个试验因素是用药后不同时间,它有 5 个水平(6h、2d、4d、9d 和 14d),而且它是个重复测量因素。因此原文采用 t 检验是错误的。

【案例释疑】 该资料试验设计类型为“具有一个重复测量两因素设计”,若资料满足参数检验的前提条件,则采用此设计定量资料方差分析。若资料不满足参数检验的前提条件,应做适当的变量变换,当变换后的资料满足参数检验的前提条件后,对变换后的资料采用此设计定量资料方差分析处理。

8.2.7 本节相关知识简要回顾与小结

在选择统计分析方法前,首先要清楚研究目的是什么,要判断资料的类型是什么,资料所取自的设计类型是什么,同时还要考察是否满足相应方法的应用条件,综合判断以后才能对统计分析方法作出合理的选择。例如,用参数检验处理定量资料时,通常对资料的正态性和方差齐性有要求。设计类型有成组设计、配对设计、随机区组设计、析因设计、正交设计、重复测量设计等。不同的设计类型对应的统计分析方法不一样,所以在应用时一定要注意。

8.3 定性资料统计分析中存在的问题

8.3.1 多重比较的校正问题

【案例 8-16】 原文目的探讨山西汉族人群原发性干燥综合征(pSS)与 HLA-DQ 等位基因的相关性,从基因水平上探索 pSS 的发病机制。方法:对 100 例 pSS 患者与 100 例健康对照进行 HLA-DQA1、HLA-DQB1 基因的分型;采用 χ^2 检验和 Fisher's 精确检验比较两组各等位基因频率的差异。结果:100 例 pSS 患者与 100 例健康对照共检出 9 种 HLA-DQA1 和 13 种 HLA-DQB1 等位基因型别。与健康对照相比较, pSS 患者中 HLA-DQA1 * 0501 和 HLA-DQB1 * 0201 等位基因频率显著增高($P < 0.05$), HLA-DQA1 * 0301/2 等位基因频率显著低于健康对照组($P < 0.05$),见表 8-30 和表 8-31。

表 8-30 pSS 患者与健康对照 HLA-DQA1 等位基因频率的比较 ($n=100$)

HLA-DQA1 等位基因	pSS 患者		健康对照		χ^2	P	RR	95% CI
	阳	性 基因频率 (%)	阳	性 基因频率 (%)				
0101	6	3.0	8	4.0	0.296	0.586	0.742	0.253 - 2.180
0102	21	10.5	28	14.0	1.14	0.286	0.721	0.394 - 1.317
0103	32	16.0	25	12.5	1.003	0.317	1.333	0.758 - 2.344
0104	15	7.5	17	8.5	0.136	0.721	0.873	0.423 - 1.800
0201	23	11.5	24	12.0	0.024	0.877	0.953	0.518 - 1.752
0301/2	26	13.0	49	24.5	8.681	0.003 Δ	0.460	0.273 - 0.777
0401	27	13.5	16	8.0	3.153	0.076	1.795	0.935 - 3.446
0501	44	22.0	24	12.0	7.087	0.008 Δ	2.068	1.203 - 3.557
0601	6	3.0	9	4.5	0.623	0.430	0.656	0.229 - 1.880

注: Δ 为 $P < 0.05$ 。

表 8-31 pSS 患者与健康对照 HLA-DQB1 等位基因频率的比较 ($n=100$)

HLA-DQB1 等位基因	pSS 患者		健康对照		χ^2	P	RR	95% CI
	阳	性 基因频率 (%)	阳	性 基因频率 (%)				
0201	57	28.5	37	18.5	5.563	0.018 ^a	1.756	1.097 - 2.812
0302	22	11.0	24	12.0	0.098	0.754	0.906	0.490 - 1.676
0301/4	32	16.0	42	21.0	1.658	0.198	0.717	0.431 - 1.192
0303	26	13.0	29	14.5	0.190	0.663	0.881	0.498 - 1.558
0401	5	2.5	9	4.5	1.184	0.276	0.544	0.179 - 1.653
0402	3	1.5	2	1.0		0.686 [#]	1.508	0.249 - 9.121
0501	4	2.0	11	5.5	3.394	0.065	0.351	0.110 - 1.120
0502	1	0.5	3	1.5		0.623 [#]	0.330	0.034 - 3.200
0503	11	5.5	5	2.5	2.344	0.126	2.270	0.774 - 6.657
0601	7	3.5	4	2.0	0.841	0.359	1.777	0.512 - 6.169
0602	5	2.5	9	4.5	1.184	0.276	0.544	0.179 - 1.653
0604	3	1.5	7	3.5	1.641	0.200	0.420	0.107 - 1.647
0603/8	24	12.0	18	9.0	0.958	0.328	1.379	0.721 - 2.629

注: #为 Fisher 精确检验; ^a $P < 0.05$ 。

【案例辨析】 原文比较了 pSS 患者与健康对照在 HLA-DQA1 和 HLA-DQB1 等位基因上的差异,原文采用了 χ^2 检验。每个等位基因都进行了一次统计学检验,多次检验会导致假阳性率迅速增大,从而使得结论不可信,因此需要对 P 值进行统计学校正。

【案例释疑】常用的多重比较校正方法是 Bonferroni 法校正 P 值, 即 $P_c = \text{原 } P \text{ 值} \times \text{比较的次数}$, $P_c \leq 0.05$ 为差异有统计学意义。两种人群间等位基因频率比较的 $P_c = \text{原 } P \text{ 值} \times \text{等位基因数}$; HLA-DQA1 的 $P_c = \text{原 } P \text{ 值} \times 9$; HLA-DQB1 的 $P_c = \text{原 } P \text{ 值} \times 13$ 。重新计算得到的 P_c 与 0.05 比较再看差异是否有统计学意义。

从表 8-32 和表 8-33 可见, 校正 P 值后仅 HLA-DQA1 * 0301/2 等位基因频率在 pSS 患者与健康对照之间差异有统计学意义 ($P < 0.05$)。HLA-DQA1 * 0501 和 HLA-DQB1 * 0201 等位基因频率在校正 P 值后两组间差异均无统计学意义 ($P > 0.05$)。

表 8-32 pSS 患者与健康对照 HLA-DQA1 等位基因频率分布比较

HLA-DQA1 等位基因	pSS 患者 ($n=100$)		健康对照 ($n=100$)		χ^2	P	P_c
	例数	频率 (%)	例数	频率 (%)			
0101	6	3.0	8	4.0	0.296	0.586	1.000
0102	21	10.5	28	14.0	1.14	0.286	1.000
0103	32	16	25	12.5	1.003	0.317	1.000
0104	15	7.4	17	8.5	0.136	0.721	1.000
0201	23	11.5	24	12.0	0.024	0.877	1.000
0301/2	26	13.0	49	24.5	8.681	0.003	0.027
0401	27	13.5	16	8.0	3.153	0.076	0.684
0501	44	22.0	24	12.0	7.087	0.008	0.072
0601	6	3.0	9	4.5	0.623	0.430	1.000
总计	200	100.0	200	100.0			

表 8-33 pSS 患者与健康对照 HLA-DQB1 等位基因频率分布比较

HLA-DQA1 等位基因	pSS 患者 ($n=100$)		健康对照 ($n=100$)		χ^2	P	P_c
	例数	频率 (%)	例数	频率 (%)			
0201	57	28.5	37	18.5	5.563	0.018	0.234
0302	22	11.0	24	12.0	0.098	0.754	1.000
0301/4	32	16.0	42	21.0	1.658	0.198	1.000
0303	26	13.0	29	14.5	0.190	0.663	1.000
0401	5	2.5	9	4.5	1.184	0.276	1.000
0402	3	1.5	2	1.0		0.686	1.000
0501	4	2.0	11	5.5	3.394	0.065	0.845
0502	1	0.5	3	1.5		0.623	1.000
0503	11	5.5	5	2.5	2.344	0.126	1.000
0601	7	3.5	4	2.0	0.841	0.359	1.000
0602	5	2.5	9	4.5	1.184	0.276	1.000
0604	3	1.5	7	3.5	1.641	0.200	1.000
0603/8	24	12.0	18	9.0	0.958	0.328	1.000
总计	200	100.0	200	100.0			

【案例 8-17】原文目的探讨维生素 D 受体 (VDR) FokI 基因多态性与系统性红斑狼疮 (SLE) 的相关性。研究对象选取 271 例 SLE 患者和 130 例健康人, 检测其 VDR FokI 基因位点及基因型分布。组间基因型及等位基因比较采用 χ^2 检验, $P < 0.05$ 为差异有统计学意义, 见表 8-34。

表 8-34 SLE 患者及正常对照人群 Fok 1 等位基因及基因型分布

		对照组 (130) <i>n</i> (%)	SLE (271) <i>n</i> (%)	χ^2	<i>P</i>	OR(95%CI)
Fok 1 等位基因	F	127(48.8)	330(60.9)	10.39	0.001	1.630(1.210-2.196)
	f	133(51.2)	212(39.1)			0.613(0.455-0.827)
Fok 1 基因型	FF	33(25.4)	116(42.8)	11.417	0.001	2.200(1.385-3.493)
	Ff	61(46.9)	98(36.2)	4.251	0.039	0.641(0.419-0.979)
	ff	36(27.7)	57(21.0)	2.187	0.139	0.695(0.429-1.127)

【案例辨析】 SLE 患者和健康人 VDR Fok I 基因型频率的比较应该属于两组构成比的比较, 首先应该整体检验两组构成是否有差异, 而原文并未给出两组整体检验的结果。一般进行整体检验构成比差异有统计学意义的话可进一步做多重比较, 但检验水准需要 Bonferroni 校正, 因为构成比合计是固定的, 所以校正数为两两比较次数-1。当检验某一基因型时, 该基因型频数不变, 其他两个基因型合并。原文多重比较时 *P* 值并未校正。

【案例释疑】 整体检验两组构成比差异有统计学意义($\chi^2 = 11.420$, $P = 0.003$)。多重比较结果见表 8-35。

表 8-35 VDR Fok I 基因型频率分布在 SLE 患者和健康人组间的比较

基因型	频率(%)		χ^2	原 <i>P</i> 值	校正 <i>P</i> 值
	对照组 (<i>n</i> =130)	SLE 患者 (<i>n</i> =271)			
FF	33(25.4)	116(42.8)	11.417	0.001	0.002
Ff	61(46.9)	98(36.2)	4.251	0.039	0.078
ff	36(27.7)	57(21.0)	2.187	0.139	0.278

整体检验: $\chi^2 = 11.420$, $P = 0.003$ 。

经过校正后, 仅 FF 基因型频率在两组间差异有统计学意义($P < 0.05$), 而 Ff 基因型校正之前 $P < 0.05$, 校正后差异无统计学意义。

8.3.2 误用一般 χ^2 检验实现趋势分析

【案例 8-18】 原文目的探讨宫颈液基细胞中转录因子 Brn-3a 表达及其早期诊断宫颈病变的临床价值。选择进行液基细胞学检查(TCT)患者 192 例, 进行 Brn-3a 蛋白免疫细胞化学染色, 并与组织病理学结果进行对照。根据 2011 年修订的 TBS 诊断系统分类, NILM 是未见上皮内病变及癌变, ASCUS 是非典型性鳞状上皮细胞, LSL 是低度鳞状上皮内病变, HSIL 是高度鳞状上皮内病变, SCC 是宫颈癌。统计方法计数资料多个样本率的比较和两个样本率之间的比较采用 χ^2 检验。 $P < 0.05$ 表示差异具有统计学意义。结果显示, 在宫颈液基细胞学诊断中, TCT 残留标本中 Brn-3a 蛋白表达的阳性率随宫颈病变严重程度加重而增高, 且各级细胞学诊断间比较差异具有统计学意义($P < 0.05$), 见表 8-36。

【案例辨析】 从原文可知 TCT 细胞学诊断 NILM、ASCUS、LSL、HSIL、SCC 的病变程度是依次增加的, 原作者目的是研究是否随着宫颈病变严重程度加重 Brn-3a 蛋白表达的阳性率有增高的趋势, 而原文作者采用了一般 χ^2 检验是错误的, 这个资料应该采用 Cochran-Armitage 线性趋势 χ^2 检验。

【案例释疑】 经趋势 χ^2 检验得 $\chi^2 = 89.071$, $P < 0.05$, 可以认为随宫颈病变严重程度的加重 Brn-3a 蛋白表达的阳性率逐渐增高。

表 8-36 宫颈液基细胞中 Brn-3a 蛋白的表达

TCT 细胞学诊断	<i>n</i>	Brn-3a 蛋白表达各级的例数		阳性率(%)
		-	+	
NILM	20	20	0	0
ASCUS	40	32	8	20.0 ^a
LSL	58	19	39	67.2
HSIL	43	4	39	90.7
SCC	31	0	31	100.0

注：^a*P* < 0.05 vs 其他组。

8.3.3 误用 χ^2 检验处理结果变量为有序变量的单向有序列联表资料

【案例 8-19】 原文目的探讨端粒酶逆转录酶(hTERT)、c-myc、Ki-67 与肝癌临床病理特征的关系以及 3 个因子之间的相互关系。方法：检测 37 例肝癌组织、5 例正常肝组织中 hTERT 和 c-myc、Ki-67 的表达，并进行对比研究。结果显示 hTERT、c-myc、Ki-67 在肝癌组织中的表达强度与年龄、性别、转移及肿瘤体积无关。三者的表达强度均随肝癌病理组织学分化程度下降而明显增强(*P* < 0.01)。统计方法计数资料用 χ^2 检验及 Fisher 精确概率法，计量资料用方差分析，相关性分析采用 Spearman 等级相关分析，*P* < 0.05 为有统计学意义，见表 8-37。

表 8-37 肝癌组织中 hTERT、c-myc、Ki-67 表达与临床病理学特征关系

Clinicopathological		<i>n</i>	hTERT 各级的例数				c-myc 各级的例数				Ki-67 各级的例数			
			-	+	++	+++	-	+	++	+++	-	+	++	+++
Age	< 60	22	3	6	10	3	1	11	10	0	0	6	8	8
	≥60	15	2	7	6	0	1	7	6	1	1	4	6	4
Sex	Man	32	4	12	13	3	2	16	14	0	1	10	11	10
	Female	5	1	1	3	0	0	2	2	1	0	0	3	2
Volume of tumor(cm)	<5	12	3	3	5	1	1	6	5	0	1	3	4	4
	≥5	25	2	10	11	2	1	12	11	1	0	7	10	8
Metastasis	Yes	17	3	5	8	1	1	7	8	1	0	5	5	4
	No	20	2	8	8	2	1	11	8	0	1	5	9	8
Histology stage	I	5	3	1	1	0	2	3	0	0	1	4	0	0
	II	17	2	12	3	0	0	14	3	0	0	6	9	2
	III	15	0	0	12	3	0	1	13	1	0	0	5	10

【案例辨析】 原文研究 hTERT、c-myc、Ki-67 三者的表达强度和肝癌病理组织学分化程度的相关关系时采用 Spearman 等级相关分析是正确的，因为它们都是等级变量。但在研究 hTERT、c-myc、Ki-67 在肝癌组织中的表达强度与年龄、性别、转移及肿瘤体积的关系时采用 χ^2 检验及 Fisher 精确概率法是错误的。此时的结局变量 hTERT、c-myc、Ki-67 三者的表达强度是等级变量，因而属于结果变量为有序变量的 2 × *C* 列联表资料，应该用秩和检验来处理，而 χ^2 检验与变量的有序性没有联系，它仅是在检验各行上的频数分布规律是否相同，而不能得出表达强度差别是否具有统计学意义的结论。

【案例释疑】 该研究目的是比较不同年龄、不同性别、不同肿瘤体积、不同转移情况下的 hTERT、c-myc、Ki-67 三者的表达强度是否有差异，因结果变量是有序变量，所以资料属于单向有序 2 × *C* 列联表资料，应采用两个独立样本比较的 Wilcoxon 秩和检验来处理，见表 8-38、表 8-39 和表 8-40。

表 8-38 肝癌组织中 hTERT 表达与临床病理学特征关系

Clinicopathological		n	hTERT 各级的例数				Z	P
			-	+	++	+++		
Age	<60	22	3	6	10	3	-1.192	0.233
	≥60	15	2	7	6	0		
Sex	Man	32	4	12	13	3	-0.024	0.981
	Female	5	1	1	3	0		
Volume of tumor	<5	12	3	3	5	1	-0.503	0.615
	≥5	25	2	10	11	2		
Metastasis	Yes	17	3	5	8	1	-0.196	0.845
	No	20	2	8	8	2		

表 8-39 肝癌组织中 c-myc 表达与临床病理学特征关系

Clinicopathological		n	hTERT 各级的例数				Z	P
			-	+	++	+++		
Age	<60	22	1	11	10	0	-0.172	0.863
	≥60	15	1	7	6	1		
Sex	Man	32	2	16	14	0	-1.090	0.276
	Female	5	0	2	2	1		
Volume of tumor	<5	12	1	6	5	0	-0.543	0.587
	≥5	25	1	12	11	1		
Metastasis	Yes	17	1	7	8	1	-0.816	0.415
	No	20	1	11	8	0		

表 8-40 肝癌组织中 Ki-67 表达与临床病理学特征关系

Clinicopathological		n	hTERT 各级的例数				Z	P
			-	+	++	+++		
Age	<60	22	0	6	8	8	-0.688	0.492
	≥60	15	1	4	6	4		
Sex	Man	32	1	10	11	10	-1.105	0.269
	Female	5	0	0	3	2		
Volume of tumor	<5	12	1	3	4	4	-0.258	0.797
	≥5	25	0	7	10	8		
Metastasis	Yes	17	0	5	5	4	-0.497	0.619
	No	20	1	5	9	8		

8.3.4 误用 χ^2 检验处理双向有序列联表资料

【案例 8-20】 原文目的探讨宫颈液基细胞中 p16INK4a 和 cyclin E 蛋白表达及其对宫颈上皮内肿瘤形成(CIN)I 级患者预后的评估价值。方法: 收集经病理组织学证实为 CIN I 级患者 86 例, 采用免疫细胞化学染色检测 p16INK4a 和 cyclin E 蛋白, 同时随访 1 年, 观察 p16INK4a 及 cyclin E 蛋白表达与 CIN I 患者病情转归的相关性。结果显示 86 例 CIN I 级患者宫颈液基细胞中 p16INK4a 阳性表达 59 例, 随访 1 年后 28 例治愈, 20 例未愈, 11 例恶化; 27 例 p16INK4a 表达阴性者中 26 例治愈, 1 例仍未愈, 均无恶化; 两者比较有统计学差异($P < 0.01$)。70 例 cyclin E 阳性表达病例中 29 例治愈, 26 例未愈, 15 例恶化; 而 16 例 cyclin E 表达阴性者均治愈。两者比较有统计学差异($P < 0.01$)。统计分析应用 SPSS 13.0 统计软件, 计数资料采用 χ^2 检验, 见表 8-41。

表 8-41 p16INK4a 及 cyclin E 蛋白的表达与 CIN I 级患者病情转归的关系

表 达 程 度		病情转归的例数		
		治 愈	未 愈	恶 化
p16INK4a 表达	-	26	1	0
	+	21	6	2
	++	7	10	5
	+++	0	4	4
cyclin E 表达	-	16	0	0
	+	29	4	1
	++	9	12	3
	+++	0	5	7

【案例辨析】 原文目的是研究 p16INK4a 及 cyclin E 蛋白表达与 CIN I 患者病情转归的相关性。从表看出，资料涉及两个多水平变量：p16INK4a(cyclin E)的表达(-、+、++、+++)和患者病情转归(治愈、未愈、恶化)，并且两个变量均具有一定顺序，因而该资料属于双向有序且属性不同的列联表资料。原作者对此定性资料进行一般 χ^2 检验是错误的，它并不能回答两个有序变量间的相关性问题。

【案例释疑】 根据研究的目的不同使用的统计分析方法也不一样。如果目的是分析 p16INK4a(cyclin E)蛋白表达与 CIN I 患者病情转归是否存在相关关系，应采用 Spearman 秩相关分析(等级相关分析)。如果研究目的是比较 p16INK4a(cyclin E)表达阳性者和表达阴性者在病情转归上是否有差异，这时结果变量是有序变量，可视为单向有序 $2 \times C$ 列联表，选用 Wilcoxon 秩和检验。表格形式见表 8-42。

表 8-42 p16INK4a 及 cyclin E 蛋白表达与 CIN I 患者病情转归的关系

蛋白表达		病情转归的例数			
		治 愈	未 愈	恶 化	合 计
p16INK4a	+	28	20	11	59
	-	26	1	0	27
cyclin E	+	29	26	15	70
	-	16	0	0	16

8.3.5 本节相关知识简要回顾与小结

定性资料统计分析方法最常用的是 χ^2 检验，但是 χ^2 检验的错误应用也是最多的。对于不同类型的列联表资料统计分析方法不同，而且一定要结合研究目的选择方法。 χ^2 检验本身也有应用条件，有校正的 χ^2 检验，有 Fisher 确切概率法作补充。除此之外还有进行趋势分析的趋势 χ^2 检验和分层分析的 CMH χ^2 检验。同时要注意多重比较的校正问题。对于等级资料，常选用秩和检验来推断等级强度的差别，若用 χ^2 检验，则只能推断各行频数分布规律是否相同。

8.4 相关与回归分析方面存在的问题

8.4.1 误用均值间差异比较处理相关性问题的

【案例 8-21】 原文目的探讨新生儿缺氧缺血性脑病(HIE)患儿脑脊液及血清中 IL-1 β 和 TNF-

α 的表达量与其脑损伤程度的关系。检测 24 例 HIE 患儿(其中轻度 5 例、中度 7 例、重度 12 例)和 13 例正常新生儿脑脊液及血清中 IL-1 β 、TNF- α 水平,并对其表达量与 HIE 的严重程度的关系进行分析。结果显示脑脊液中 IL-1 β 及 TNF- α 的表达水平与临床脑损伤程度具有相关性。在 HIE 轻度、中度及重度患儿脑脊液中 IL-1 β 的表达水平随着脑损伤的程度的加重其表达呈上升趋势。TNF- α 的表达水平呈上升趋势但没有 IL-1 β 的趋势明显,见图 8-1。

【案例辨析】 原文目的是研究随着临床脑损伤程度的加重 IL-1 β 及 TNF- α 的表达水平是否呈上升趋势,原作者在这里未作统计学检验,仅通过比较图中均值的大小就作出判断是很不可靠的。

【案例释疑】 要通过相关分析来确定是否有相关关系,因临床脑损伤程度是等级变量,IL-1 β 或 TNF- α 的表达水平是连续性变量,所以要采用 Spearman 秩相关分析研究两变量之间的相关性。如果资料是双变量正态分布的数据可用 Pearson 相关分析。

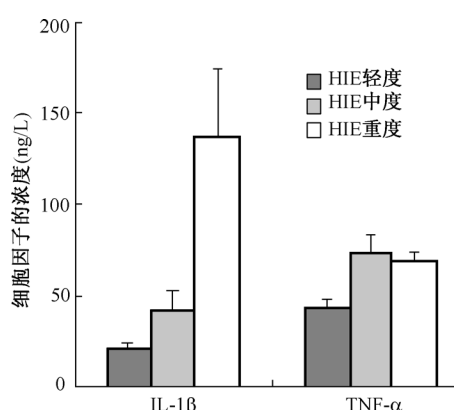


图 8-1 脑脊液中 IL-1 β 和 TNF- α 的表达随 HIE 轻度、中度及重度损伤程度的加重而升高

8.4.2 Cox 回归分析方面存在的问题

【案例 8-22】 原文目的检测胶质瘤组织中转化生长因子 $\beta 1$ (TGF- $\beta 1$) 表达,并分析其与 CD4⁺ FOXP3⁺ 调节性 T 细胞(Treg)浸润的关系,探讨 Treg 浸润的临床意义。方法:检测 135 例胶质瘤组织、15 例正常脑组织中 TGF- $\beta 1$ 、Treg(CD4⁺ FOXP3⁺) 的表达。多因素 Cox 回归分析显示 TGF- $\beta 1$ 表达、Treg 细胞数量不是影响胶质瘤患者预后的独立危险因素($P > 0.05$),而患者的年龄、性别、KPS 评分、病理级别是影响预后的独立危险因素($P < 0.05$)。生存分析采用 Kaplan-Meier 及 log-rank 检验分析方法。所有检验结果以 $P < 0.05$ (双侧)认定为有统计学意义,见表 8-43。

表 8-43 胶质瘤患者总生存时间影响因素的 Cox 回归模型分析

变 量	危 险 比 率	变 量	危 险 比 率
年龄	1.765 (1.130 - 2.757) ^a	病理级别	1.951 (1.186 - 3.208) ^b
性别	0.586 (0.356 - 0.964) ^a	TGF- $\beta 1$ 表达	1.519 (0.946 - 2.439)
KPS 评分	0.275 (0.162 - 0.466) ^b	Treg 阳性细胞数	1.284 (0.817 - 2.017)

注: ^a $P < 0.05$ vs 对照组; ^b $P < 0.01$ vs 对照组。

【案例辨析】 首先原文缺乏必要的变量赋值说明,因为结果的解释必需要知道变量是怎么赋值的。原文得到的 Cox 回归分析结果是极其不规范的,缺乏很多必要的结果,例如回归系数估计值 b 及其标准误 S_b 、Wald χ^2 值和具体的 P 值大小等。同时原文也缺乏对变量筛选方法以及引入和剔除检验水准的说明。

【案例释疑】 首先应给出胶质瘤可能的影响因素及其赋值,见表 8-44。

表 8-44 胶质瘤可能的影响因素与赋值

因 素	变 量 名	赋 值 说 明
年龄	X_1	$\leq 39 = 1, > 39 = 2$
性别	X_2	女 = 0, 男 = 1

续表		
因 素	变 量 名	赋 值 说 明
KPS 评分	X_3	<70 分 = 1, ≥ 70 分 = 2
病理级别	X_4	低级别 = 1, 高级别 = 2
TGF- β 1 表达	X_5	低表达 = 1, 高表达 = 2
Treg 阳性细胞数	X_6	(个)
生存时间	t	(月)
生存结局	Y	删失 = 0, 死亡 = 1

以生存时间、生存结局为因变量, 以 $X_1 \sim X_6$ 为协变量, 进行 Cox 回归模型分析, 同时要写明变量筛选的方法、引入检验水准 $\alpha_{\text{入}}$ 和剔除检验水准 $\alpha_{\text{出}}$ 。保留在模型中的变量的 P 值都应小于剔除检验水准。得到的 Cox 回归模型拟合结果见表 8-45。

表 8-45 胶质瘤影响因素的 Cox 回归分析结果

选 入 变 量	回归系数估计值 b	标准误 S_b	Wald χ^2	P 值	RR 的估计值	RR 95%CI
...
...
...

8.4.3 本节相关知识简要回顾与小结

Cox 回归模型是目前进行多因素生存分析的主要方法, 它以生存时间、生存结局为因变量, 探讨可能影响生存时间的有关因素。当协变量个数较多时可进行变量的筛选, 筛选变量的方法有前进法、后退法和逐步回归法, 同样要求设置引入自变量的检验水准 $\alpha_{\text{入}}$ 和剔除自变量的检验水准 $\alpha_{\text{出}}$ 。对于无序分类变量要进行哑变量处理, 它们作为一个整体同时进出模型。但 Cox 回归模型要求满足比例风险假定(即协变量对生存率的影响不随时间的改变而改变)这个前提条件, 只有该条件得到满足, Cox 比例风险回归模型结果才有效。

参 考 文 献

[1] 孙振球. 医学统计学(第 3 版). 北京: 人民卫生出版社, 2010.
[2] 方积乾. 卫生统计学(第 6 版). 北京: 人民卫生出版社, 2008.
[3] 胡良平. 统计学三型理论在试验设计中的应用. 北京: 人民军医出版社, 2006.
[4] 胡良平. 生物医学科研课题的研究设计与统计分析——错误案例辨析与释疑. 北京: 军事医学科学出版社, 2008.
[5] 余松林, 向惠云. 重复测量资料分析方法与 SAS 程序. 北京: 科学出版社, 2004.

第 9 章 血液学统计设计和数据分析

目前已有部分文章及专著报道了有关医学科研论文中常见的统计设计问题和统计分析错误，但对《中华血液学杂志》中存在的统计问题尚未有专门的报道。本章选取《中华血液学杂志》2007—2011 年编辑出版的、刊载于论著以及短篇论著专栏的论文，归纳总结了该杂志中的科研论文所用到的统计软件、研究内容和统计分析方法，同时对所涉及的科研设计和统计分析及统计结果表达中常见的错误进行辨析与释疑。

本章内容包含以下几个方面：(1) 杂志中所用的统计软件和研究内容及统计分析方法归类；(2) 科研设计的常见错误并释疑；(3) 统计分析方法应用方面的常见错误并释疑；(4) 统计结果报告中常见问题并释疑。

9.1 杂志近 5 年论文分析

从表 9-1 到表 9-4 可知：该杂志中常用的统计分析方法为 t 检验和 χ^2 检验，分别占到 22.45% 和 17.09%；常用的统计软件为 SAS 和 SPSS，分别占到 7.16% 和 89.40%；研究淋巴细胞和浆细胞疾病及造血干细胞疾病的文章最多，分别占到 29.79% 和 39.61%。

表 9-1 2007 – 2011《中华血液学杂志》常用统计分析方法汇总结果

统计方法名称	例数 (%)					
	2007 年	2008 年	2009 年	2010 年	2011 年	合 计
t 检验	68(17.66)	63(20.59)	70(26.51)	66(22.45)	68(27.98)	335(22.45)
方差分析	45(11.69)	37(12.09)	34(12.88)	42(14.29)	33(13.58)	191(12.80)
卡方检验	38(9.87)	50(16.34)	54(20.45)	69(23.47)	44(18.11)	255(17.09)
非参数检验	116(30.13)	8(2.61)	11(4.17)	17(5.78)	18(7.41)	170(11.39)
Kaplan-Meier	23(5.97)	42(13.73)	40(15.15)	47(15.99)	30(12.35)	182(12.20)
log-rank 检验	11(2.86)	26(8.50)	25(9.47)	31(10.54)	22(9.05)	115(7.71)
Cox 回归	11(2.86)	15(4.90)	12(4.55)	11(3.74)	10(4.12)	59(3.95)
Logistic 回归	60(15.58)	4(1.31)	4(1.51)	4(1.36)	9(3.70)	81(5.43)
直线相关	8(2.08)	50(16.34)	6(2.27)	1(0.34)	5(2.06)	70(4.69)
Spearman 秩相关	5(1.30)	11(3.59)	8(3.03)	6(2.04)	4(1.65)	34(2.28)

注：部分文章所用统计分析方法超过 1 种。

表 9-2 2007 – 2011《中华血液学杂志》统计软件应用汇总结果

		例数 (%)					
		2007 年	2008 年	2009 年	2010 年	2011 年	合 计
是否用统计软件	用	116(57.14)	134(67.34)	124(63.59)	144(70.94)	152(81.28)	670(67.88)
	不用	87(42.86)	65(32.66)	71(36.41)	59(29.06)	35(18.72)	317(32.12)
	合计	203(100.00)	199(100.00)	195(100.00)	203(100.00)	187(100.00)	987(100.00)
统计软件	SAS	16(13.79)	11(8.21)	10(8.06)	5(3.47)	6(3.95)	48(7.16)
	SPSS	90(77.59)	122(91.04)	110(88.71)	135(93.75)	142(93.42)	599(89.40)

续表

		例数(%)					
		2007 年	2008 年	2009 年	2010 年	2011 年	合 计
统计软件	Excel	2 (1.72)	0(0.00)	0 (0.00)	0(0.00)	1(0.66)	3(0.45)
	Stata	6(5.17)	0(0.00)	2(1.61)	3(2.08)	1 (0.66)	12(1.79)
	其他软件	2(1.72)	1(0.75)	2(1.61)	1(0.7)	2(1.32)	8(1.19)
	合计	116(100.00)	134(100.00)	124(100.00)	144(100.00)	152(100.00)	670(100.00)

表 9-3 2007 -2011《中华血液学杂志》统计方法应用汇总结果

是否有统计方法	例数(%)					
	2007 年	2008 年	2009 年	2010 年	2011 年	合 计
有	144(70.94)	143(71.86)	159(81.54)	160(78.82)	131(70.05)	737(74.67)
无	59(29.06)	56(28.14)	36(18.46)	43(21.18)	56(29.95)	250(25.33)
合计	203(100.00)	199(100.00)	195(100.00)	203(100.00)	187(100.00)	987(100.00)

表 9-4 2007 -2011《中华血液学杂志》内容归类汇总结果

血液病分类	例数(%)					
	2007 年	2008 年	2009 年	2010 年	2011 年	合 计
红细胞疾病	6(2.96)	5(2.51)	4(2.05)	9(4.43)	6(3.21)	30(3.04)
粒细胞疾病	0(0.00)	3(1.51)	2(1.03)	0(0.00)	1(0.53)	6(0.61)
单核细胞和巨噬细胞疾病	2(0.99)	4(2.01)	3(1.54)	3(1.48)	7(3.74)	19(1.93)
淋巴细胞和浆细胞疾病	60(29.56)	58(29.15)	62(31.79)	61(30.05)	53(28.34)	294(29.79)
造血干细胞疾病	54(26.6)	102(51.26)	101(51.79)	70(34.48)	64(34.22)	391(39.61)
脾功能亢进	1(0.49)	0(0.00)	0(0.00)	1(0.49)	0(0.00)	2(0.20)
出血性及血栓性疾病	11(5.42)	21(10.55)	19(9.74)	22(10.84)	21(11.23)	94(9.52)
其他	69(33.99)	6(3.02)	4(2.05)	37(18.23)	35(18.72)	151(15.30)
合计	203(100.00)	199(100.00)	195(100.00)	203(100.00)	187(100.00)	987(100.00)

注：血液系统疾病分类参考陆再英和钟南山主编《内科学》，人民卫生出版社。

9.2 科研设计中的常见问题及案例辨析

9.2.1 回顾性描述资料及未进行统计分析直接下结论

【案例 9-1】 探讨 22 号染色体三体 (+22) 在诊断 inv(16) 急性髓系白血病(AML) 中的价值。采用红、绿荧光直接标记的双色断裂点分离的基因探针 CBFβ 对 18 例存在 +22 克隆异常的 AML 患者进行间期荧光原位杂交(FISH)检测，并与 R 显带常规细胞遗传学(CC)检测结果进行比较分析。应用多重荧光原位杂交(M-FISH)技术检测同时伴有 +22 和 inv(16)的 AML 患者。结果 18 例存在 +22 的 AML 患者中，CC 分析未发现 inv(16)阳性，而 FISH 检测发现 11 例 inv(16)阳性、1 例 del(16)(q22)。这 11 例阳性患者中，9 例 CC 分析为单纯 +22 异常，1 例伴有 +8，1 例患者为 del(16)(q22)，而 FISH 检测为 inv(16)。M-FISH 检测 4 例同时伴有 +22 和 inv(16)的患者，除 +22 外，未发现其他异常。结论 +22 是预测 inv(16)AML 的重要标志，具有潜在的诊断 inv(16)AML 的价值。全文未用任何统计分析方法。

【案例 9-2】 从细胞周期角度探讨豆类凝集素 FRIL(Flt3 Receptor- Interacting Lectin)体外维持造血干/祖细胞静息的分子机制。以添加 FRIL、Flt3 配体(FL)和不添加因子的培养基分别培养脐血 CD34⁺细胞，采用 RT-PCR 和 Western blot 法分别从 mRNA 和蛋白水平检测细胞周期相关分子的

表达。结果:新分离的 CD34⁺ 细胞中未检测到 G₀/G₁ 期相关 Cyclin 和 CDK 蛋白的表达,培养 3d 时 FRIL 组 CyclinD3、CDK6 蛋白相对表达量(分别为 483 ± 63、553 ± 39)低于两个对照组(FL 组: 2437 ± 52、3209 ± 98;空白对照组: 914 ± 105、1497 ± 55);培养 3d 时 FRIL 组 P27 蛋白相对表达量(0.312 ± 0.030)低于对照组(FL 组: 0.787 ± 0.024;空白对照组: 0.616 ± 0.029),但表达较高的 P53 蛋白(FRIL 组、FL 组、空白对照组分别为 4.476 ± 0.159、0.581 ± 0.099、2.167 ± 0.114)。FRIL 组细胞周期正向调节因子的 mRNA 相对表达量低于或相当于 FL 组和空白对照组。结论 FRIL 通过抑制参与造血细胞周期调控的 CyclinD3 和 CDK6 的表达,延迟了造血干/祖细胞的细胞周期,P27 在 FRIL 抑制造血干/祖细胞分化中发挥了重要作用,P53 也参与了 FRIL 对造血干/祖细胞的维持作用。全文未用任何统计分析方法。

【案例 9-3】 研究蛋白 C(PC)基因 C64W、F139V 和 K150 缺失突变(K150d)致 PC 缺陷症的分子机制。结果 PC C64W 未从转染细胞内分泌且在细胞内逐渐降解,PC F139V 仅部分从细胞内分泌,大部分未分泌并在细胞内逐渐降解,PC K150d 突变体蛋白的分泌未受突变的明显影响。real-time PCR 结果表明,与野生型 PC mRNA 相比,突变体 PC mRNA 不降低;蛋白降解抑制试验显示,PC C64W 和 PC F139V 通过蛋白酶体途径进行细胞内降解。Endo H 酶切和转染细胞荧光染色显示,PC C64W 和 PC F139V 主要位于前高尔基体。结论分泌障碍和细胞内降解是 PC C64W 和 PC F139V 导致 PC 缺陷症的原因,PC K150d 对突变体蛋白的分泌无明显影响。全文未用任何统计分析方法。

【案例 9-4】 评价骨髓增生异常综合征(MDS)WHO 标准和 MDS 最低诊断标准对骨髓原始细胞比例不增高(<0.050)的 MDS 患者的诊断分型。回顾性收集 1988 年至 2005 年诊断的骨髓原始细胞比例不增高的 210 例 MDS 患者,重新分析细胞形态并采集相关试验室检查资料,用 WHO 标准和 MDS 最低诊断标准对这些患者进行诊断分型。结果:按 WHO 分型标准进行诊断分型,其中 RA 5 例,RARS 7 例,难治性血细胞减少伴有多系发育异常(RCMD)76 例,RCMD 和环状铁粒幼红细胞(RCMD-RS)9 例,MDS 不能分类(MDS-U)35 例,5q-综合征 3 例,无法分型 75 例。WHO 标准不能分型的患者中,16 例骨髓二系或三系发育异常,而血细胞仅一系减少的患者随访半年后 9 例出现二系以上血细胞减少,可归为 RCMD;对 44 例骨髓仅红系发育异常,血细胞二系减少或全血细胞减少无法进行分型患者,重新进行形态学分析,将那些尚未述及的细胞发育异常形态改变纳入判定该系细胞发育异常细胞计数范围后,27 例患者粒系形态异常细胞 ≥ 0.100,可归入 RCMD 或 RCMD-RS;15 例三系均无细胞发育异常患者,采用 MDS 最低诊断标准诊断为 MDS 5 例,高度疑似 MDS(HS-MDS)5 例,意义未定的特发性血细胞减少(ICUS)5 例。结论 MDS 中只有很少一部分患者是真正意义上的 RA;关于 MDS 的细胞发育异常形态学异常的范畴和量的界定值仍值得进一步探讨;MDS 最低诊断标准是对 WHO 标准的很好的细化和补充。全文未用任何统计分析方法。

【案例 9-5】 提高对 T 大颗粒淋巴细胞白血病(T-LGLL)的认识。回顾性分析我院 1999 年 6 月至 2007 年 7 月间确诊的 27 例 T-LGLL 患者临床及试验室检查特征。结果:贫血相关症状最为突出,51.9% 患者脾脏轻中度肿大,14.8% 患者肝脏轻度肿大。未有合并类风湿性关节炎者。外周血中性粒细胞 < 1.5 × 10⁹/L 者 19 例,占 70.4%, < 0.5 × 10⁹/L 者 5 例,占 18.5%。贫血占 88.9%,中位血红蛋白 58 g/L,合并纯红细胞再生障碍者 18 例,占 66.67%。外周血大颗粒淋巴细胞(LGL)绝对值中位数 1.45 × 10⁹/L,21 例(77.8%)患者 LGL 数 ≤ 2.0 × 10⁹/L,22 例(81.5%)患者 LGL 免疫表型为 CD3⁺CD8⁺CD57⁺CD56⁻。免疫抑制治疗总有效率 91.3%,完全血液学缓解率为 65.2%。结论 T-LGLL 起病及进展缓慢,主要表现为血细胞减少和脾脏肿大,常合并纯红细胞再生障碍。多数患者外周血 LGL 计数 ≤ 2.0 × 10⁹/L,免疫表型以 CD3⁺CD8⁺CD57⁺为主,对免疫抑制治疗反应良好。全文未用任何统计分析方法。

【案例辨析】 以上五个案例的共性问题是原文作者未对资料进行统计分析就得出结论，即仅对资料进行描述得出结论，缺乏科学的统计分析，故所得结论不可靠。

【案例释疑】 应根据研究目的、设计类型、资料类型和资料所满足的前提条件选择合理的统计分析方法进行统计推断，再得出结论。

9.2.2 组间均衡性差

【案例 9-6】 研究 CD28、CTLA-4 在再生障碍性贫血(AA)免疫发病机制中的可能作用。23 例发病期 AA 患者均为 2005 年 7 月至 2006 年 3 月在我院就诊患者，其中重型 AA(SAA)6 例，慢性 AA(CAA)17 例，男 15 例，女 8 例，中位年龄 33(8~66)岁。10 例恢复期 AA 患者，其中 SAA 2 例，CAA 8 例。男 5 例，女 5 例，中位年龄 25(11~50)岁，所有恢复患者脱离输血，血常规：中性粒细胞绝对值(ANC)≥1.0×10⁹/L，BPC≥20×10⁹/L；骨髓增生活跃，粒系、红系达正常比例，巨核细胞≥8 个/片。正常对照组 15 名，男 4 名，女 11 名，中位年龄 35(22~51)岁。数值比较采用 *t* 检验。

【案例辨析】 此研究违背了均衡原则。原文作者收治了 23 例发病期 AA 患者、10 例恢复期 AA 患者和正常对照组 15 名，但原文作者未对三组研究对象的基本情况进行分析，以确保三组研究对象在某些重要的非试验因素方面具有均衡性，从文中可知三组研究对象在性别、年龄等重要非试验因素方面的差异比较大，不均衡，故所得结论不可靠。

【案例释疑】 研究者在临床试验设计阶段必须考虑性别、年龄等重要非试验因素对研究结果的影响，在设计阶段使各组研究对象在“样本例数、男女比例、各年龄段例数等”方面都接近，也就是设置的对照组除研究因素所取的水平与试验组不同外，其他对研究结果有影响的因素(非试验因素)应尽量一致，所得结论才具有说服力。

9.2.3 样本例数太少

【案例 9-7】 探讨套细胞淋巴瘤(Mantle Cell Lymphoma, MCL)患者的临床特征、不同治疗方案的疗效及预后因素。回顾性分析 2000 年 1 月至 2010 年 3 月在四川大学华西医院住院治疗的 37 例 MCL 患者的临床特征及不同治疗方案的疗效，并进行相关预后因素分析。统计学处理：采用 SPSS 13.0 统计软件进行统计学分析。正态分布或近似正态分布的计量资料以 $\bar{x} \pm s$ 表示，非正态分布计量资料以中位数表示，组间率的比较采用 Fisher 确切概率法。部分结果如下。

治疗反应：37 例患者中，方案一的患者有效率 46.15%(13 例中 6 例)，方案二的患者有效率 42.86%(7 例中 3 例)，方案三的患者有效率 92.31%(13 例中 12 例)，使用方案四的 3 例患者均有效，方案五的 1 例患者无效。不同方案间疗效差异经 Fisher 确切概率法检验，方案三疗效优于方案一及方案二($P < 0.05$)，方案一与方案二、方案四与其他方案间疗效差异均没有统计学意义(P 值均 > 0.05)，见表 9-5。

表 9-5 不同治疗方案的疗效比较[例(%)]

组 别	例 数	完全缓解	部分缓解	稳 定	进 展	总有效率(%)
方案一	13	2(15.38)	4(30.77)	6(46.15)	1(7.69)	46.15*
方案二	7	1(14.29)	2(28.57)	4(57.14)	0	42.86*
方案三	13	4(30.77)	8(61.54)	1(7.69)	0	92.31
方案四	3	0	3(100.00)	0	0	100.00
方案五	1	0	0	1(100.00)	0	0

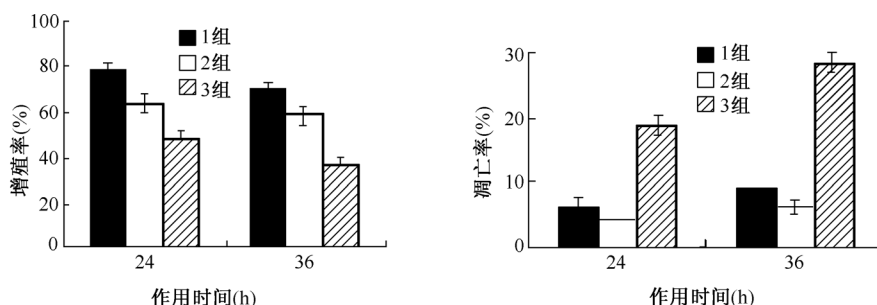
注：* 与方案三比较， $P < 0.05$ 。

【案例辨析】 第一, 37 例 MCL 患者分五组用不同治疗方案研究其疗效, 但是原文作者未对五组病人的基本情况进行分析, 故不能保证五组患者间的一般情况具有均衡性; 第二, 在表 9-5 中, 计算有效率的分子太小, 分母太小时, 不仅代表性不够且计算的相对数不稳定, 如果样本含量小于 20, 不宜计算相对数。

【案例释疑】 第一, 对表 9-5 中五组患者的一般情况进行统计分析, 判断资料是否符合均衡原则; 第二, 表 9-5 中五个总有效率的分子均小于 20, 其中三组小于等于 7, 故只需给出各组的总例数和实际发生某现象的例数; 第三, 应大力提倡前瞻性临床试验研究, 即在确定研究课题之后, 应结合基本常识、专业知识和统计学知识, 制订出科学完善的临床试验设计方案(应正确考虑伦理道德问题、盲法设计问题、严格掌握三要素和四原则、选择合适的设计类型, 注重研究过程中的质量控制), 在此方案的指导下, 开展临床试验研究, 并选用合理的统计分析方法处理所获得的临床资料, 方可得出经得起时间和实践检验的结论。

9.2.4 对照不全

【案例 9-8】 硼替佐米联合三氧化二砷诱导 NB4 细胞凋亡的机制研究。部分结果见图 9-1。



(a) 单药组及联合组对 NB4 细胞生长抑制作用的比较

(b) 单药组及联合组对 NB4 细胞诱导凋亡作用的比较

1 组: As_2O_3 ($0.5 \mu\text{mol/L}$); 2 组: 硼替佐米 (10 nmol/L);

3 组: As_2O_3 ($0.5 \mu\text{mol/L}$) + 硼替佐米 (10 nmol/L)

图 9-1 硼替佐米联合 As_2O_3 对 NB4 细胞生长及凋亡的影响

【案例辨析】 该资料除时间外还涉及两个因素: As_2O_3 和硼替佐米用否, 二者都有两个水平: 用和不用。这两个因素水平的全面组合应当有 4 个组, 原文作者少设置了一个对照组, 即两样都不用, 犯了“对照不全”的错误。该资料用表格的形式(表 9-6)给出, NB4 细胞凋亡率的表格与表 9-6 类似。

表 9-6 单药组及联合组对 NB4 细胞生长抑制作用的试验结果

As_2O_3 用否	硼替佐米用否	NB4 细胞生存率 ($\bar{x} \pm s$)		As_2O_3 用否	硼替佐米用否	NB4 细胞生存率 ($\bar{x} \pm s$)	
		* : 24	36			* : 24	36
否	否	X	X	是	否	X	X
	是	X	X		是	X	X

注: “X”代表相应的数据; *: 代表“作用时间(h)”。

【案例释疑】 从表 9-6 可知, 该试验设计是一个 2×2 析因设计, 若同时考虑时间, 就是“具有一个重复测量的三因素设计”。原文作者应加上一组对照, 即 As_2O_3 与硼替佐米都不用。

9.2.5 本节相关知识简要回顾

统计设计包括调查设计、试验设计和临床试验设计, 其中试验设计尤为重要。试验设计需重点

掌握“三要素”、“四原则”和“设计类型”。三要素包括受试对象、影响因素和试验效应；四原则包括重复、对照、随机和均衡原则；常见的试验设计类型有单因素设计(包括单组设计、配对设计、成组设计和单因素多水平设计)、析因设计和重复测量设计。前瞻性的完全随机对照试验所得的结论最可靠，而回顾性描述和分析资料不易满足试验设计的四项原则，故所得结论不可靠。

9.3 定量资料统计分析中的常见问题及案例辨析

9.3.1 忽略参数检验的前提条件

【案例 9-9】 总结多发性骨髓瘤(MM)患者骨病(MBD)的临床特点。回顾性分析 205 例 MM 患者 MBD 的临床特点，比较不同 MBD 分级患者的临床疗效及 MBD 分级与预后的关系，见表 9-7。统计学处理：不同 MBD 分级患者的血液生化指标比较采用非参数统计秩和检验中的 Kruskal-Wallis H 检验。

表 9-7 不同 MBD 分级患者临床特点比较

组 别	骨髓中浆细胞 比例 (%)	CD138 ⁺ CD38 ⁺ 细胞比例	血清 IL-6 (ng/L)	血磷 (mmol/L)	血 LDH (U/L)	血 ALB (g/L)
无溶骨性损害 (0-1 级)组	0.345 ± 0.036	3.6 ± 0.4	92.8 ± 5.4	1.8 ± 0.1	214.8 ± 24.4	32.7 ± 1.1
有溶骨性损害 (2-3 级)组	0.343 ± 0.029	13.4 ± 2.6	103.6 ± 7.1	1.7 ± 0.1	211.3 ± 22.6	32.5 ± 0.7
并发病理性骨折 (4 级)组	0.407 ± 0.024	26.8 ± 4.2	134.5 ± 9.5	1.8 ± 1.1	200.4 ± 21.5	32.6 ± 0.7

【案例辨析】 该资料分组变量为等级变量，结果变量为定量变量，为单因素三水平设计定量资料，原文作者选用非参数统计中的秩和检验分析该资料，但文中并未提到正态性和方差齐性检验，选用参数检验还是非参数检验应视结果变量是否满足参数检验前提条件(如正态性和方差齐性)，符合参数检验条件的资料而使用非参数检验方法分析，会降低检验效能。

【案例释疑】 该单因素三水平设计定量资料如果满足单因素方差分析的应用条件(独立性、正态性和方差齐性)，可使用单因素方差分析；如果资料不满足单因素方差分析的前提条件，可进行适当的变量变换，当变换后的资料满足单因素方差分析的前提条件，仍可采用该法分析，否则，只能选择适合的非参数检验方法，如秩和检验。

【案例 9-10】 通过测定不同类型淋巴瘤患者血浆 TNF-α、TNF-β 水平的变化，探讨血浆 TNF-α、TNF-β 水平的变化与临床预后指标的关系。统计学处理：数据用 $\bar{x} \pm s$ 表示，两独立样本均数比较采用 t 检验或 t' 检验，多个均数比较采用方差分析，见表 9-8。

【案例辨析】 在应用 t 检验和单因素方差分析时，资料需要满足独立性、正态性和方差齐性。本例部分数据的标准差大于均数，说明这些数据可能不服从正态分布，因此，第一，不能用 $\bar{x} \pm s$ 描述非正态分布的定量资料的分布特征；第二，资料不满足 t 检验和方差分析的前提条件，应用 t 检验和方差分析处理该资料不合适。

【案例释疑】 第一，描述呈偏态分布的定量资料的集中趋势和离散程度应选用中位数和四分位数间距；第二，找到满足上述前提条件的适当变量变换方法后，再使用相应的参数检验，否则选用适合分析偏态分布资料的非参数检验如秩和检验进行分析；第三，对于这类情况，可以在表格的下方备注各指标使用的统计学方法。

表 9-8 淋巴瘤患者血浆 TNF- α 、TNF- β 水平与临床预后指标的关系 (ng/L, $\bar{x} \pm s$)

临床因素		例数	TNF- α	P 值	TNF- β	P 值
性别	男	78	47 \pm 63		270 \pm 320	
	女	54	31 \pm 23	0.074	236 \pm 265	0.517
年龄(岁)	>60	55	48 \pm 72		288 \pm 372	
	<60	77	41 \pm 65	0.565	214 \pm 140	0.156
民族	汉族	90	44 \pm 60		236 \pm 242	
	少数民族	46	34 \pm 25	0.280	296 \pm 388	0.273
IPI 指数	0 ~ 1	16	35 \pm 19		322 \pm 457	
	2 ~ 3	41	38 \pm 27		256 \pm 288	
	4 ~ 5	19	78 \pm 118	0.046	412 \pm 476	0.121
Ann Arbor 分期	I ~ II	36	36 \pm 25		273 \pm 312	
	III ~ IV	73	48 \pm 66	0.295	247 \pm 274	0.652
LDH(IU/L)	≥ 300	14	55 \pm 38		267 \pm 414	
	<300	108	41 \pm 58	0.392	267 \pm 298	0.999
$\beta 2$ 微球蛋白(mg/L)	≤ 3	41	39 \pm 25		290 \pm 399	
	>3	40	45 \pm 86	0.686	312 \pm 344	0.788
C 反应蛋白(mg/L)	<10	37	34 \pm 24		357 \pm 445	
	>10	31	54 \pm 97	0.234	207 \pm 101	0.053
B 症状	无	78	31 \pm 24		231 \pm 278	
	有	49	57 \pm 77	0.005	293 \pm 341	0.262

9.3.2 误用 t 检验处理单因素多水平设计定量资料

【案例 9-11】 观察辐射后骨髓基质细胞(MSC)高迁移率族蛋白 B1(High Mobility Group Box 1, HMGB1)的释放,研究 HMGB1 对人脐血造血干细胞(HSC)增殖分化的影响。集落形成试验:将新鲜分离的人脐血单个核细胞以 1×10^5 /孔($300\mu\text{l}$)接种于 12 孔细胞培养板中,每孔加入上述半固体培养基 3ml,设 4 组,分别为不加 HMGB1 对照组, HMGB1 终浓度分别为 5ng/ml、50ng/ml 和 100ng/ml 三个试验组,每组设一平行孔。置于 5% CO_2 、37℃、饱和湿度的孵箱中培养 14 d 后,倒置显微镜下分别计数红系爆式集落形成单位(BFU-E)及粒-单核细胞集落形成单位(CFU-GM)和总集落的生成数。统计学处理:用 SPSS 11.5 统计软件进行分析,结果以 $\bar{x} \pm s$ 表示,采用 t 检验进行分析。部分结果见表 9-9。

表 9-9 HMGB1 对脐血造血干祖细胞集落形成的影响($\bar{x} \pm s$, $n=8$)

HMGB1 浓度 (ng/ml)	集落数			HMGB1 浓度 (ng/ml)	集落数		
	BFU-E	CFU-GM	总 CFU		BFU-E	CFU-GM	总 CFU
0	37 \pm 18	48 \pm 7	88 \pm 24	50	67 \pm 28	93 \pm 12	165 \pm 44
5	54 \pm 21	71 \pm 9	130 \pm 35	100	71 \pm 29	96 \pm 14	173 \pm 47

注: HMGB1 各浓度组与 0 ng/ml 组比较, P 值均 <0.05 。

【案例辨析】 误用 t 检验处理宜用单因素方差分析处理的资料。表 9-9 所对应的试验设计类型是一个单因素四水平设计定量资料,原作者选用 t 检验分析该资料不妥当, t 检验通常只用于分析单组设计、配对设计或成组设计定量资料,采用 t 检验分析该资料,会增大犯 I 型错误的概率。

【案例释疑】 对于上述四组定量资料的每一个指标,若资料满足参数检验的前提条件,应选用单因素四水平设计定量资料的方差分析,如果上述三个指标间有联系,选用单因素四水平设计定量资料的多元方差分析;若资料不满足参数检验的前提条件,应寻找合适的变量变换方法,当变换后的资料满足参数检验的前提条件时,对变换后的资料采用单因素四水平设计定量资料的方差分

析处理，否则选用单因素四水平设计定量资料的 Kruskal-Wallis 秩和检验。当四组间集落数差别有统计学意义时，需进行两两比较，两两比较时不宜直接采用 t 检验，因为会增大犯 I 型错误的概率，应选择专用的两两比较的方法。

9.3.3 误用析因设计定量资料方差分析和成组设计定量资料 t 检验处理
具有一个重复测量两因素设计定量资料

【案例 9-12】 探讨骨髓间充质干细胞(MSC)对半相合骨髓移植小鼠免疫功能的影响及机制。经 $8\text{ Gy}^{60}\text{Co}\gamma$ 射线照射后，BALB/c($H-2^d$) 雌性小鼠分为两组：MSC 组 and 对照组。观察两组移植后不同时间外周血中的淋巴细胞亚群，ConA 和 LPS 刺激的 T、B 淋巴细胞增殖，供受者及第三者混合淋巴细胞反应(MLR)，供者骨髓及脾细胞在受者胸腺、骨髓和脾脏的植入率，供者 MSC 在受者体内分布，急、慢性 GVHD 的发生率和生存分析。统计学处理：数据以 $\bar{x} \pm s$ 表示。SAS9.0 软件对数据进行析因设计、成组 t 检验和生存分析。部分结果见表 9-10。

表 9-10 骨髓移植后两组小鼠外周血 T 细胞亚群检测结果 ($n=20, \bar{x} \pm s$)

组 别	CD3 (%)				CD4 (%)			
	7 天	14 天	30 天	90 天	7 天	14 天	30 天	90 天
MSC 组	8.21 \pm 1.21	31.38 \pm 4.62 *	35.43 \pm 4.83 *	32.31 \pm 3.11	14.52 \pm 3.11	42.31 \pm 4.91	48.59 \pm 5.52	38.06 \pm 7.26
对照组	4.41 \pm 0.52	8.68 \pm 2.61	20.12 \pm 1.87	34.57 \pm 4.25	47.51 \pm 7.43	38.10 \pm 2.85	45.31 \pm 6.74	40.26 \pm 7.12

组别	CD8 (%)				CD4/CD8			
	7 天	14 天	30 天	90 天	7 天	14 天	30 天	90 天
MSC 组	4.62 \pm 1.56	33.52 \pm 6.33	18.52 \pm 3.46	52.26 \pm 5.72	3.14 \pm 0.56	1.26 \pm 0.23 *	3.62 \pm 0.56 *	0.83 \pm 0.36
对照组	7.53 \pm 1.71	55.95 \pm 6.56	42.32 \pm 5.45	45.06 \pm 6.52	6.33 \pm 1.23	0.68 \pm 0.12	1.07 \pm 0.14	0.84 \pm 0.56

注：与对照组比较，* $P < 0.05$ 。

【案例辨析】 第一，原文结果部分没给出检验统计量；第二，原文未说明哪部分用 t 检验，哪部分用析因设计定量资料方差分析。本资料有两个试验因素，一个因素为“组别”，具有两个水平(MSC 组和对照组)；另一个因素为“观测时间”，具有四个水平(7、14、30、90 天)。每只小鼠在 7、14、30、90 天 4 个时间点上被重复测量了同一指标的数值，4 个时间点测得的数据具有相关性，因而该资料属于具有一个重复测量两因素设计定量资料。原作者无论用析因设计定量资料方差分析还是成组设计定量资料 t 检验处理数据都不合适。

【案例释疑】 对于 CD3(%)、CD4(%)、CD8(%) 和 CD4/CD8 四个指标，如果资料满足参数检验的应用条件，可以采用具有一个重复测量两因素设计定量资料的方差分析处理，如果这 4 个指标间在专业上有一定联系，可采用具有一个重复测量两因素设计定量资料四元方差分析处理；如果资料不满足参数检验的应用条件，采用适当的变量变换方法后，再使用相应的参数检验，否则选用适合的非参数检验，如秩和检验。

【案例 9-13】 探讨高毒力株幽门螺杆菌感染与慢性原发免疫性血小板减少症的关系及幽门螺杆菌根除治疗的临床疗效。62 例 Hp 感染的 CITP 患者中 CagA 蛋白抗体检测结果显示：高毒力株 Hp 感染者[CagA 蛋白抗体阳性(CagA^+)]26 例，低毒力株 Hp 感染患者(CagA^-)36 例。继续将 CagA^+ 患者随机分为治疗组和对照组，每组各 13 例； CagA^- 患者亦随机分为治疗组与对照组，每组各 18 例。治疗方案：对照组采用肾上腺糖皮质激素单独治疗方案；治疗组在对照组肾上腺糖皮质激素治疗方案基础上加用根除 Hp 治疗。统计学处理：采用 SPSS 17.0 统计分析软件，血小板计数组间比较，采用 t 检验。部分结果见表 9-11 与表 9-12。

表 9-11 CagA⁺ CITP 患者治疗组与对照组治疗前后血小板计数变化 ($\times 10^9/L$)

组 别	例 数	治 疗 前	治 疗 后
治疗组	13	15.38 \pm 9.92	31.69 \pm 28.36 ^a
对照组	13	18.69 \pm 10.91	18.08 \pm 16.02 ^{bc}

注:与治疗前比较, ^a $t=2.63$, $P<0.05$; ^b $t=0.189$, $P>0.05$; 与对照组治疗后比较, ^c $t=2.61$, $P<0.05$ 。

表 9-12 CagA⁻ CITP 患者治疗组与对照组治疗前后血小板计数变化 ($\times 10^9/L$)

组 别	例 数	治 疗 前	治 疗 后
治疗组	18	18.22 \pm 9.76	19.28 \pm 14.13 ^a
对照组	18	20.50 \pm 10.25	20.50 \pm 10.25 ^{bc}

注:与治疗前比较, ^a $t=0.42$, $P<0.05$; ^b $t=0.22$, $P>0.05$ 。与对照组治疗后比较, ^c $t=0.30$, $P>0.05$ 。

【案例辨析】 原文作者所采用的统计分析方法不正确。本资料有两个试验因素,一个因素为“组别”,其有两个水平(治疗组和对照组);另一个因素为“治疗时间”,其有两个水平(治疗前和治疗后),因为治疗前和治疗后两个时间点的数据具有相关性,因而该资料属于具有一个重复测量两因素设计定量资料。 t 检验仅适合分析单因素设计且因素的水平数最多为2的定量资料。原作者采用 t 检验进行统计分析,会增大犯I型错误的概率,结论不可靠。

【案例释疑】 如果资料满足参数检验的前提条件,本例可以选用具有一个重复测量两因素设计定量资料方差分析处理;如果资料不满足参数检验的应用条件,采用适当的变量变换方法后,再使用相应的参数检验,否则选用适合的非参数检验,如秩和检验。

9.3.4 未用统计方法分析具有一个重复测量三因素设计定量资料

【案例 9-14】 探讨尿多酸肽(CDA-II)单独和联合环腺苷酸(cAMP)对维甲酸耐药的急性早幼粒细胞白血病(APL)细胞的作用。未用任何统计分析方法。部分结果如下:NB4-R2 细胞经 1 g/L CDA-II 单独处理 48 h 和 72 h 时, Bcl-2 表达呈阳性的细胞比例分别由(92.0 \pm 0.6)% 下降至(40.5 \pm 2.2)% 和(26.8 \pm 9.2)%; 当 1 g/L CDA-II 联合 100 μ mol/L cAMP 共同处理 NB4-R2 细胞时, 48 h 和 72 h 时 Bcl-2 阳性细胞比例分别下降至(15.1 \pm 4.8)% 和(7.3 \pm 2.9)%; 100 μ mol/L cAMP 单独可使 NB4-R2 细胞的 Bcl-2 阳性率由(92.0 \pm 0.6)% 下降至(75.3 \pm 2.0)%, 且几乎不会因作用时间的延长而继续下降,见表 9-13。

表 9-13 CDA-II 和(或)cAMP 处理对 NB4-R2 细胞内 Bcl-2 蛋白表达水平的影响 ($\bar{x} \pm s$, $n=3$)

组 别	Bcl-2 阳性细胞百分率 (%)		组 别	Bcl-2 阳性细胞百分率 (%)	
	48 h	72 h		48 h	72 h
对照组	92.0 \pm 0.6	92.0 \pm 0.6	1 g/L CDA-II 组	40.5 \pm 2.2	26.8 \pm 9.2
100 μ mol/L cAMP 组	75.3 \pm 2.0	72.9 \pm 6.4	cAMP + CDA-II 组	15.1 \pm 4.8	7.3 \pm 2.9

【案例辨析】 原文作者未进行统计分析,仅根据对试验数据的描述就直接下结论。

【案例释疑】 除时间外这四组涉及两个因素:cAMP 剂量(0 与 100) μ mol/L 和 CDA-II 剂量(0 与 1)g/L,二者都有两个水平:其中 0 剂量可视为“不用此药物”、另一个剂量就是以特定剂量来使用此药。这两个因素水平的全面组合就是表中组别对应的四组,同时考虑时间,该资料所取自的设计类型就是“具有一个重复测量三因素设计”,若资料满足参数检验的前提条件,则采用与该设计对应的定量资料方差分析;若资料不满足参数检验的前提条件,应进行适当的变量变换,当变换后的资料满足参数检验的前提条件后,对变换后的资料采用与该设计对应的定量资料方差分析处理。该资料设计类型由表 9-14 给出。

表 9-14 CDA- II 和 (或) cAMP 处理对 NB4 - R2 细胞内 Bcl-2 蛋白表达水平的影响 ($\bar{x} \pm s$, $n=3$)

100 μ mol/L cAMP 用否	1 g/L CDA- II 用否	Bcl-2 阳性细胞百分率 (%)	
		作用时间 (h): 48	72
否	否	92.0 \pm 0.6	92.0 \pm 0.6
	是	40.5 \pm 2.2	26.8 \pm 9.2
是	否	75.3 \pm 2.0	72.9 \pm 6.4
	是	15.1 \pm 4.8	7.3 \pm 2.9

9.3.5 未用统计方法分析两因素析因设计定量资料

【案例 9-15】 探讨尿多酸肽 (CDA-II) 单独和联合环腺苷酸 (cAMP) 对维甲酸耐药的急性早幼粒细胞白血病 (APL) 细胞的作用。未用任何统计分析方法。部分结果如下: 当 NB4 细胞经 1g/L CDA- II 单独处理 72 h 时, 细胞表面分化抗原 CD11c 和 NBT 的阳性率都有一定程度的升高; 而当采用 100 μ mol/L cAMP 与 CDA- II 联合处理时, NB4 细胞的 CD11c 表达和 NBT 阳性率则会明显提高。此外, 我们还检测了细胞内凋亡抑制蛋白 Bcl-2 的表达情况, 发现无论是 CDA- II 单独还是联合 cAMP 都会显著抑制 NB4 细胞内 Bcl-2 的表达 (表 9-15), 从形态上看, 此时细胞表现出部分分化的现象, 并伴有凋亡发生。这些现象充分表明 CDA- II 单独或联合 cAMP 能够使维甲酸敏感的 NB4 细胞发生部分分化, 并促进凋亡。

表 9-15 CDA- II 和 (或) cAMP 处理对 NB4 细胞表面分化抗原 CD11c、NBT 阳性率以及细胞内 Bcl-2 蛋白表达水平的影响 ($\bar{x} \pm s$, $n=3$)

组 别	CD11c 阳性细胞百分率 (%)	NBT (A ₅₄₀)	Bcl-2 阳性细胞百分率 (%)
对照组	19.6 \pm 0.5	0.062 \pm 0.021	73.7 \pm 2.1
100 μ mol/L cAMP 组	69.1 \pm 10.1	0.066 \pm 0.032	13.6 \pm 9.1
1 g/L CDA- II 组	29.9 \pm 5.9	0.136 \pm 0.005	27.8 \pm 8.2
cAMP + CDA- II 组	72.2 \pm 12.0	0.350 \pm 0.011	10.5 \pm 2.3

【案例辨析】 原作者未进行统计分析, 仅根据试验数据直接下结论。

【案例释疑】 表 9-15 涉及两个试验因素: cAMP 的剂量 (0 与 100) μ mol/L 和 CDA- II 的剂量 (0 与 1) g/L, 二者都有两个水平: 其中 0 剂量可视为“不用此药物”, 另一个剂量就是以特定剂量来使用此药。表 9-15 未反映出资料所对应的试验设计类型, 将该表修改为表 9-16 的形式。

表 9-16 CDA- II 和 (或) cAMP 处理对 NB4 细胞表面分化抗原 CD11c、NBT 阳性率以及细胞内 Bcl-2 蛋白表达水平的影响 ($\bar{x} \pm s$, $n=3$)

100 μ mol/L cAMP 用否	1 g/L CDA- II 用否	CD11c 阳性细胞百分率 (%)	NBT (A ₅₄₀)	Bcl-2 阳性细胞百分率 (%)
否	否	19.6 \pm 0.5	0.062 \pm 0.021	73.7 \pm 2.1
	是	69.1 \pm 10.1	0.066 \pm 0.032	13.6 \pm 9.1
是	否	29.9 \pm 5.9	0.136 \pm 0.005	27.8 \pm 8.2
	是	72.2 \pm 12.0	0.350 \pm 0.011	10.5 \pm 2.3

由表 9-16 可知, 该资料的试验设计类型为 2×2 析因设计。对于该资料三项统计指标中的每一项指标, 若满足参数检验的前提条件, 都应采用两因素析因设计一元定量资料方差分析处理, 如果 CD11c 阳性细胞百分率 (%)、NBT (A₅₄₀) 和 Bcl-2 阳性细胞百分率 (%) 三个指标间在专业上有一定联系, 还可采用两因素析因设计定量资料多元方差分析处理; 若资料不满足参数检验的前提条件, 可以对资料先进行某种变量变换, 若经变量变换后的资料满足参数检验的前提条件, 仍可采用前述

的参数检验方法分析资料。

9.3.6 误用单因素设计定量资料方差分析处理两因素析因设计定量资料

【案例9-16】 研究去甲斑蝥素(NCTD)联合阿霉素(ADR)对多发性骨髓瘤细胞增殖、凋亡的影响及其机制。统计学处理:各试验独立重复3次,用SPSS 16.0软件进行统计学处理,结果均以 $\bar{x} \pm s$ 表示,组间比较采用方差分析,两两比较采用LSD检验。部分结果见表9-17。

表9-17 NCTD联合ADR对U266细胞survivin、Bcl-2和Bax表达的影响($\bar{x} \pm s$, $n=3$)

组别	survivin	Bcl-2	Bax
空白对照组	1.20 ± 0.08	0.67 ± 0.09	0.08 ± 0.01
NCTD单用组(10.0 μmol/L)	0.58 ± 0.02 ^a	0.47 ± 0.06 ^a	0.29 ± 0.04 ^a
ADR单用组(0.25 μmol/L)	0.31 ± 0.05 ^a	0.23 ± 0.05 ^a	0.46 ± 0.06 ^a
联合组(ADR 0.25 μmol/L + NCTD 10.0 μmol/L)	0.03 ± 0.02 ^b	0.05 ± 0.02 ^b	0.62 ± 0.08 ^b

注:与空白对照组比较,^a $P < 0.05$;与ADR单用组比较,^b $P < 0.05$ 。

【案例辨析】 该试验涉及两个因素:ADR的剂量(0与0.25) μmol/L和NCTD的剂量(0与10.0) μmol/L,二者都有两个水平:其中0剂量可视作“不用此药物”,另一个剂量就是以特定剂量来使用此药。这两个因素水平的全面组合就是表9-18中对应的四组,原文作者误认为组别下的四组是一个因素的4个不同水平,使用了单因素四水平设计定量资料的方差分析处理资料,忽略了两个因素之间的交互作用,犯了以单因素设计取代多因素设计的错误。

【案例释疑】 若该定量资料满足参数检验的前提条件,采用两因素析因设计定量资料方差分析,如果所测的各项指标之间存在联系,则应使用两因素析因设计定量资料多元方差分析;若该定量资料不满足参数检验的前提条件,应寻找合适的变量变换方法,当变换后的资料满足参数检验的前提条件时,对变换后的资料采用两因素析因设计定量资料方差分析处理。

表9-18 NCTD联合ADR对U266细胞survivin、Bcl-2和Bax表达的影响($\bar{x} \pm s$, $n=3$)

ADR用否	NCTD用否	survivin	Bcl-2	Bax
否	否	1.20 ± 0.08	0.67 ± 0.09	0.08 ± 0.01
	是	0.58 ± 0.02 ^a	0.47 ± 0.06 ^a	0.29 ± 0.04 ^a
是	否	0.31 ± 0.05 ^a	0.23 ± 0.05 ^a	0.46 ± 0.06 ^a
	是	0.03 ± 0.02 ^b	0.05 ± 0.02 ^b	0.62 ± 0.08 ^b

注:与空白对照组比较,^a $P < 0.05$;与ADR单用组比较,^b $P < 0.05$ 。

9.3.7 本节相关知识简要回顾

进行定量资料统计分析前需判断资料所取自的设计类型及是否满足该设计类型的应用条件,如参数检验中两总体均值比较的 t 检验要求定量资料满足独立性、正态性和方差齐性三个前提条件,若定量资料不满足该前提条件,应寻找合适的变量变换方法,当变换后的资料满足参数检验的前提条件,用参数检验法分析资料;不满足参数检验的前提条件而盲目套用参数检验法,降低了结论可信度。定量资料常用的设计类型有成组设计、配对设计、单因素多水平设计(水平数大于等于3)、析因设计、重复测量设计等。人们一般易误用多次单因素设计分析方法处理相应的多因素设计资料。分析重复测量资料尤需谨慎,当受试对象接受某种处理后,对同一观察单位的某变量在不同时间点或不同部位进行多次测量所得的数据,为重复测量资料,但有时“时间”因素并非全是重复测量因素,如对不同的受试对象在不同时间点进行测量所得的数据,该资料不同时间点间的数据不具有相关性,故并非重复测量资料。

9.4 定性资料统计分析中的常见问题及案例辨析

9.4.1 计算率时分母过小不合适

【案例 9-17】 160 例骨髓增生异常综合征患者 JAK2 V617F 突变的检测及临床意义。统计学处理：计数资料的比较采用 χ^2 检验，计量资料的比较采用秩和检验。部分结果见表 9-19。

表 9-19 MDS 患者 JAK2 V617F 突变型与野生型的临床特征 ($\bar{x} \pm s$)

组 别	例 数	性别(例)		年龄(例)		核型异常 (例)	脾肿大 (例)
		男	女	<60 岁	≥60 岁		
突变型	8	5	3	1	7	2	3
野生型	152	78	74	117	35	51	8
P 值		0.799		0.794		0.908	0.005

突变型患者脾肿大者 3 例，野生型患者脾肿大者 8 例，突变型脾肿大发生率高于野生型($P = 0.005$)。突变型患者核型异常发生率为 25.0% (8 例中 2 例)，野生型患者为 33.6% (152 例中 51 例)，差异无统计学意义($P = 0.908$)。

【案例辨析】 表 9-19 中计算突变型脾肿大发生率的分母是 8，分母太小，显得过分夸张，计算的发生率不稳定。

【案例释疑】 第一，根据研究目的计算相应的样本含量，重新收集资料并重新分析；第二，该例突变型脾肿大组样本含量小于 20，故只需给出各组总例数和实际发生某现象的例数，没有必要计算发生率。

【案例 9-18】 探讨套细胞淋巴瘤(Mantle Cell Lymphoma, MCL)患者的临床特征、不同治疗方案的疗效及预后因素。回顾性分析 2000 年 1 月至 2010 年 3 月在四川大学华西医院住院治疗的 37 例 MCL 患者的临床特征及不同治疗方案的疗效，并进行相关预后因素分析。统计学处理：采用 SPSS 13.0 统计软件进行统计学分析。正态分布或近似正态分布的计量资料以 $\bar{x} \pm s$ 表示，非正态分布计量资料以中位数表示，组间率的比较采用 Fisher 确切概率法。部分结果如下：治疗反应：37 例患者中，方案一的患者有效率 46.15% (13 例中 6 例)，方案二的患者有效率 42.86% (7 例中 3 例)，方案三的患者有效率 92.31% (13 例中 12 例)，使用方案四的 3 例患者均有效，方案五的 1 例患者无效。不同方案间疗效差异经 Fisher 确切概率法检验，方案三疗效优于方案一及方案二($P < 0.05$)，方案一与方案二、方案四与其他方案间疗效差异均没有统计学意义(P 值均 > 0.05)，结果见表 9-20。

表 9-20 不同治疗方案的疗效比较[例(%)]

组 别	例 数	完全 缓解	部分 缓解	稳 定	进 展	总有效率(%)
方案一	13	2(15.38)	4(30.77)	6(46.15)	1(7.69)	46.15 ^a
方案二	7	1(14.29)	2(28.57)	4(57.14)	0	42.86 ^a
方案三	13	4(30.77)	8(61.54)	1(7.69)	0	92.31
方案四	3	0	3(100.00)	0	0	100.00
方案五	1	0	0	1(100.00)	0	0

注：^a与方案三比较， $P < 0.05$ 。

【案例辨析】 表 9-20 中计算总有效率的分子是 13、7、13、3 和 1，分母太小。

【案例释疑】 第一，根据研究目的计算相应的样本含量，重新收集资料并重新分析；第二，该例五组样本含量均小于 20，故只需给出各组总例数和实际发生某现象的例数，没有必要计算有效率。

9.4.2 误用 Fisher 确切概率法处理单向有序的列联表资料

【案例 9-19】 同例 9-18(表 9-21)。

表 9-21 不同治疗方案的疗效比较[例(%)]

组别	例数	完全缓解	部分缓解	稳定	进展	总有效率(%)
方案一	13	2(15.38)	4(30.77)	6(46.15)	1(7.69)	46.15 ^a
方案二	7	1(14.29)	2(28.57)	4(57.14)	0	42.86 ^a
方案三	13	4(30.77)	8(61.54)	1(7.69)	0	92.31
方案四	3	0	3(100.00)	0	0	100.00
方案五	1	0	0	1(100.00)	0	0

注：^a与方案三比较， $P < 0.05$ 。

【案例辨析】 表 9-21 中结果变量疗效是有序变量，因而属结果变量为有序变量的单向有序二维列联表资料。原文作者将结果变量疗效分为完全缓解、部分缓解、稳定和进展 4 个等级，但计算有效率时将不同水平的疗效进行了合并，即将完全缓解和部分缓解合并，稳定和进展合并，组间率的比较采用 Fisher 确切概率法，这样做损失了数据的有序性信息。

【案例释疑】 宜采用秩和检验分析该资料。

9.4.3 误用 χ^2 检验分析结果变量为有序变量的列联表资料

【案例 9-20】 硼替佐米为主方案治疗多发性骨髓瘤患者的临床研究。统计学处理：采用 SPSS 17.0 统计软件进行分析。计数资料和疗效的比较采用卡方检验。部分结果如下：总体疗效：39 例患者总反应率(CR + nCR + PR + MR)为 92.3%。22 例初治患者中 21 例(95.5%)对治疗有反应。17 例复发、难治患者中 15 例(88.2%)对治疗有反应。初治组患者的 CR 率明显高于复发/难治组(40.9%对 5.9%)，差异有统计学意义($P = 0.024$)，但两组患者的总反应率差异无统计学意义($P = 0.570$)，结果见表 9-22。

【案例辨析】 表 9-22 中结果变量疗效是多值有序变量，因而属结果变量为有序变量的单向有序二维列联表资料。原文作者将结果变量疗效分为 CR、nCR、PR、MR 和 NC + PD 五个等级，又将 CR、nCR、PR、MR 四个水平进行合并，使用 χ^2 检验进行分析，这样做损失了数据的有序性信息。

【案例释疑】 宜采用秩和检验分析该资料(表 9-22)。

表 9-22 39 例多发性骨髓瘤患者总体疗效[例数(%)]

组别	例数	CR	nCR	PR	MR	NC + PD
初治组	22	9(40.9)	6(27.3)	6(27.3)	0	1(4.5)
复发/难治组	17	1(5.9)	2(11.8)	11(64.7)	1(5.9)	2(11.8)
合计	39	10(25.6)	8(20.5)	17(43.6)	1(2.6)	3(7.7)

注：CR：完全缓解；nCR：接近完全缓解；PR：部分缓解；MR：轻微反应；NC：无变化；PD：疾病进展；初治组与复发/难治组 CR 率、CR + nCR 率比较， P 值均 < 0.05 ；两组总反应率(CR + nCR + PR + MR)比较， $P = 0.570$ 。

9.4.4 误用 χ^2 检验取代多重 logistic 回归分析

【案例 9-21】 探讨慢性淋巴细胞白血病(CLL)患者脂蛋白酯酶(LPL)和血清胸苷激酶(TK)水平及其与其他 CLL 预后因素的相关性。统计学处理： χ^2 检验分析 TK1 水平与其他预后因素的关系。部分结果如下：血清 TK1 浓度与 CLL 其他预后因素的关系：39 例 CLL 患者血清 TK1 浓度与其他预后因素的相关性见表 9-23。在外周血淋巴细胞绝对计数(ALC)明显增高组($\geq 50 \times 10^9/L$)、血清乳

酸脱氢酶(LDH)高水平组、无 IgVH 基因突变组和 ZAP-70 蛋白高表达组($\geq 20\%$)的患者中,血清 TK1 浓度分别明显高于 ALC 无明显增高组($< 50 \times 10^9/L$)($P=0.018$)、血清 LDH 水平正常组($P=0.018$)、具有 IgVH 基因突变组($P=0.030$)和 ZAP-70 蛋白低表达组($< 20\%$)($P=0.038$)的患者。血清 TK1 浓度与 CLL 患者性别、年龄、Binet 分期、CD38 表达水平和遗传学异常无明显相关性($P>0.05$)。

表 9-23 血清 TK1 浓度与其他预后因素的关系

因 素		例 数	TK1 阳性例数(%)	P 值
性别	男	28	7(25.0)	0.501
	女	11	2(18.2)	
年龄	≥ 60 岁	25	4(16.0)	0.157
	< 60 岁	14	5(35.7)	
Binet 分期	A + B	31	31(19.4)	0.260
	C	8	3(37.5)	
外周血 ALC	$\geq 50 \times 10^9/L$	6	4(66.7)	0.018
	$< 50 \times 10^9/L$	33	5(15.2)	
血清 LDH 水平	增高	9	5(55.6)	0.018
	正常	30	4(13.3)	
IgVH 基因突变率	$\geq 2\%$	26	4(15.4)	0.030
	$< 2\%$	9	5(55.6)	
CD38 表达率	$< 30\%$	29	5(17.2)	0.150
	$\geq 30\%$	10	4(40.0)	
ZAP-70 表达率	$< 20\%$	25	3(12.0)	0.038
	$\geq 20\%$	14	6(42.9)	
分子遗传学	del(17p13)或 del(11q22)	8	3(37.5)	0.336
	单独伴有 del(13q14)	11	2(18.2)	
	其他	20	4(20.0)	

【案例辨析】 原作者误用了 χ^2 检验处理多因素设计定性资料。资料中影响血清 TK1 浓度的因素有“性别”、“年龄”、“Binet 分期”等,共 9 个因素,属于多因素分析问题,原文作者将该多因素研究资料拆分为 9 个单因素资料分别分析,采用 χ^2 检验考察某个自变量对结果的影响时无法排除其他自变量对结果的影响,所得的结论不可靠。

【案例释疑】 此处结果变量是“血清 TK1 浓度高与低”,属于结果变量为二值变量的多因素研究资料,可以采用二分类反应变量的非条件 logistic 回归分析处理此资料。

9.4.5 本节相关知识简要回顾

常用相对数(率、构成比和相对比)描述定性资料的数据特征,相对数的意义和计算方法需重点掌握,同时需注意计算相对数时分母不宜太小以及正确区分率与构成比的意义和应用,切勿以构成比代替率分析资料。常用 χ^2 检验分析结果变量为二值变量和/或多值名义变量(如癌细胞的类型 A、B、C、D 四型)定性资料,需掌握各种类型资料的 χ^2 检验方法及适用条件;对于结果变量为等级资料的定性资料,常选用秩和检验推断各处理组之间在平均等级强度上的差别是否具有统计学意义;若用 χ^2 检验,则只能推断各组之间的构成比上的差别是否具有统计学意义。如果定量资料不满足参数检验的前提条件,也可采用秩和检验进行分析。

9.5 相关回归分析中的常见问题及案例辨析

9.5.1 性质不同的数据合并进行相关分析

【案例9-22】 探讨不同的治疗方案[BD(硼替佐米+地塞米松)、TD(沙利度胺+地塞米松)]对多发性骨髓瘤(MM)骨病的影响。2006年11月至2008年9月,分别应用硼替佐米和沙利度胺联合地塞米松治疗的初治及难治复发MM患者共40例。统计学处理:应用SPSS 10.0统计软件进行分析,BD和TD方案治疗前后指标的比较采用Wilcoxon方法检验。各参数之间的相关性采用Spearman相关系数检验。部分结果:骨代谢指标及其调节因子的相关性:DKK-1、sRANKL、TRACP-5b三组数据在40例患者治疗前后均有显著的相关性($r > 0.5$)。

【案例辨析】 原文作者把应用硼替佐米和沙利度胺联合地塞米松治疗的初治及难治复发MM患者共40例且分别接受两种治疗方案的相关数据合并在一起进行Spearman秩相关分析,就是把性质不同的数据简单合并并计算相关系数,而实际上某些指标在两组间有差异,因而原文作者忽略了数据的同质性。

【案例释疑】 应分组进行相关分析,计算各组的相关系数,以保证各组内数据具有同质性。如要计算总的相关系数,可先比较各组的相关系数,如果差别无统计学意义,可合并数据,得出总的相关系数;如果差别有统计学意义,则不能将性质不同的各组数据合在一起进行计算。

9.5.2 相关分析未画散点图及相关系数过小

【案例9-23】 用流式细胞术、RT-PCR方法测定ITP患儿骨髓单个核细胞中CD40和CD40L及其mRNA的表达情况,研究其在急、慢性ITP患儿体内表达及其与血小板相关抗体PAIgG产生的关系,为阐明ITP的发病机制和寻找新的免疫治疗方法提供试验依据。统计学处理:用直线回归法进行相关性分析。部分结果:CD40、CD40L与PAIgG的相关性分析:CD40、CD40L与PAIgG呈正相关(r 值分别为0.741、0.826, $P < 0.05$)。

【案例9-24】 检测了32例MM患者骨髓单个核细胞(MNC)培养上清液的OPN水平及21例MM患者TRACP-5b,旨在探讨OPN在MM发病中的意义。统计学处理:采用SPSS 11.0软件包进行直线相关分析。部分结果:OPN与恶性浆细胞数、血 β_2 -MG水平、TRACP-5b的相关性分析:OPN水平与血 β_2 -MG(检测22例)、恶性浆细胞数(检测24例)、TRACP-5b(检测21例)均呈正相关,其 r 值分别为0.62、0.74、0.59, P 值分别为0.002、0、0.038。

【案例9-25】 了解骨髓单个核细胞Coombs试验阳性血细胞减少(CBCPC)患者补体水平的变化,探讨补体在CBCPC患者骨髓造血细胞破坏机制中的作用。统计学处理:采用Pearson相关分析。部分结果:骨髓补体C5b-9水平与CD34⁺细胞膜IgG结合率、CD34⁺细胞膜IgM结合率均呈显著正相关($r = 0.593$, $P = 0.000$; $r = 0.326$, $P = 0.049$)。

【案例9-26】 研究 β -榄香烯增强阿克拉霉素对白血病细胞系HL-60细胞的诱导凋亡作用,探讨 β -榄香烯联合阿克拉霉素抗白血病的作用机制。统计学处理:相关性分析采用直线相关分析。部分结果:当0.10 μ g/ml的阿克拉霉素与ELE联合作用时,明显抑制PGE2的分泌(与对照组比较, $P < 0.05$),且抑制效应与ELE浓度呈正相关($r = 0.787$, $P < 0.05$)。

【案例9-27】 研究骨髓增生异常综合征(MDS)患者骨髓T细胞极化状况及其与骨髓上清液中促凋亡因子TNF- α 、INF- γ 水平及造血细胞凋亡的关系。统计学处理:所有数据采用SPSS 11.0软件包进行统计学处理。两参数相关分析采用Pearson相关分析。部分结果:MDS骨髓上清液INF- γ 和TNF- α 浓度均与骨髓Th1细胞百分比呈正相关(r 值分别为0.38和0.39, P 值均小于0.05)。

【案例辨析】 以上五例的原文作者没有画散点图就进行相关分析。在分析两个定量指标之间的相关性时，应根据散点图的特征决定是否进行相关与回归分析或建立合适的曲线回归方程。若散点图中各散点没有明显的曲线趋势，且在与横轴或纵轴不平行的一条不太宽的带内随机分布，说明有线性相关的关系，就可以进行直线相关分析，即计算相关系数 r ，并且对相关系数进行假设检验。需注意相关系数的密切程度应由决定系数 (R^2) 的数值大小来度量，如例 9-27 分析得 MDS 骨髓上清液 $\text{INF-}\gamma$ 和 $\text{TNF-}\alpha$ 浓度均与骨髓 Th1 细胞百分比呈正相关， r 值分别为 0.38 和 0.39， P 值均小于 0.05，但 R^2 分别为 0.14 和 0.15，即决定系数 R^2 小于 0.5，说明两者的相关无较大的实用价值。

【案例释疑】 对有意义的单组设计二元定量资料进行相关分析时，先画出散点图，如果两变量的散点图有线性相关的关系，再计算相关系数，并且对相关系数进行检验假设，如果有意义且决定系数 R^2 大于 0.5 才能说明两变量之间的相关关系具有较大的实用价值；如果散点图显示出曲线变化关系，则应进行相应的曲线回归分析。

9.5.3 用单因素分析取代多因素分析

【案例 9-28】 血清 TK1 浓度与 CLL 其他预后因素的关系：39 例 CLL 患者血清 TK1 浓度与其他预后因素的相关性(表 9-24)。

表 9-24 血清 TK1 浓度高低与其他预后因素的关系

因 素		例 数	TK1 阳性例数(%)	P 值
性别	男	28	7(25.0)	0.501
	女	11	2(18.2)	
年龄	≥ 60 岁	25	4(16.0)	0.157
	< 60 岁	14	5(35.7)	
Binet 分期	A + B	31	31(19.4)	0.260
	C	8	3(37.5)	
外周血 ALC	$\geq 50 \times 10^9/\text{L}$	6	4(66.7)	0.018
	$< 50 \times 10^9/\text{L}$	33	5(15.2)	
血清 LDH 水平	增高	9	5(55.6)	0.018
	正常	30	4(13.3)	
IgVH 基因突变率	$\geq 2\%$	26	4(15.4)	0.030
	$< 2\%$	9	5(55.6)	
CD38 表达率	$< 30\%$	29	5(17.2)	0.150
	$\geq 30\%$	10	4(40.0)	
ZAP-70 表达率	$< 20\%$	25	3(12.0)	0.038
	$\geq 20\%$	14	6(42.9)	
分子遗传学	del(17p13)或 del(11q22)	8	3(37.5)	0.336
	单独伴有 del(13q14)	11	2(18.2)	
	其他	20	4(20.0)	

【案例 9-29】 CMV 感染发生危险因素分析情况：单因素分析 CMV 感染发生危险因素表明，年龄 < 20 岁、血缘关系 HLA 不合移植和非血缘关系移植、预处理中使用 ATG、II - IV 度 aGVHD 对于移植后 CMV 感染发生具有统计学意义(表 9-25)。

【案例辨析】 表 9-24 和表 9-25 两资料都属于多因素设计定性资料。原文作者误将一个多因素资料拆分成一系列单因素资料来处理，即用 χ^2 检验分析多因素设计定性资料，采用 χ^2 检验考察某个自变量对结果的影响时，会忽略其他自变量对结果的影响，所得结论不可靠。

表 9-25 CMV 感染单因素分析

因 素		例 数	CMV 感染例数	+ 100dCMV 感染发生率(%)	<i>P</i> 值
受者性别	男	223	91	40. 8	0. 893
	女	95	38	32. 6	
年龄(岁)	<20	65	36	55. 4	0. 006
	≥20	253	93	36. 7	
移植前疾病状态	CR/CP	294	118	40. 1	0. 585
	NR/RP/AP/BP	24	11	45. 8	
HLA 配型	同胞组合	160	28	17. 5	0. 000
	血缘关系不合	127	84	66. 1	
	非血缘关系	31	17	56. 7	
干细胞来源	BM + PBSC	250	101	40. 4	0. 913
	PBSC	62	26	41. 9	
	BM	6	2	33. 3	
ATG 应用	是	162	101	62. 3	0. 000
	否	156	28	17. 9	
大蒜素预防	是	239	98	41. 0	0. 782
	否	79	31	39. 2	
急性 GVHD	0 ~ I 度	227	69	30. 4	0. 000
	Ⅱ ~ Ⅳ度	91	60	65. 9	

【案例释疑】 表 9-24 和表 9-25 两资料的结果变量是二值变量,应采用非条件 logistic 回归分析进行处理,同时至少使用三种筛选变量的方法进行变量筛选,并根据方程评价标准确定一个最优的 logistic 回归方程,据此得出专业结论。

9.5.4 筛选变量的方法有误

【案例 9-30】 回顾性分析了我院收治的 115 例 MM 患者发病时 ALC 与化疗有效率、无病生存(DFS)和总生存(OS)时间的关系,探讨 ALC 作为 MM 患者辅助性预后判断指标的可行性。统计学处理:经单因素分析有意义的变量再进入 Cox 模型进行多因素分析,从而排除变量之间的干扰因素,确立影响疗效和生存预后的独立因素。

【案例辨析】 原文作者犯了两个错误,第一,筛选变量的策略错误,即仅根据单因素分析中有统计学意义(即 $P < 0.05$)的因素建立 Cox 回归方程不妥,因为经单因素分析无统计学意义(即 $P > 0.05$)的因素也许进行多因素分析时由于其他因素的存在而变得有统计学意义;第二,通常应选用三种及以上(如前进法、后退法和逐步回归法等)筛选自变量的方法建立 Cox 回归模型,并借助一些评价方法(如若有多个较好的 Cox 回归模型时,以残差平方和较小且模型中所含的自变量的个数较少者为最佳,模型中各偏回归系数估计值的正负号符合专业意义等),确定一个最合适的 Cox 回归模型。

【案例释疑】 正确的做法是应选用不少于三种筛选自变量的方法来建立 Cox 回归方程,并借助一些评价方法,确定一个最合适的 Cox 回归模型,据此得出专业结论。

【案例 9-31】 探讨急性髓系白血病中枢神经系统复发的高危因素。统计学分析:应用 t 检验进行组间均数比较,生存率的比较采用 log-rank 检验,分别采用 logistic 回归模型和 Cox 回归模型进行单因素和多因素预后风险分析。分析结果见表 9-26。

【案例辨析】 原文作者采取不筛选自变量的方法建立 logistic 回归模型,所以表 9-26 最后一列有 $P > 0.05$ 自变量,而且也没有对所建立的回归方程进行评价,所以以上建立的模型不是最合适的。需注意此种不筛选变量的方法应用于构建多重线性回归方程也是错误的。

表 9-26 M₃患者 CNS 复发高危因素分析

		例数	CNS 复发例数	P 值(单因素)	P 值(多因素)
性别	男	42	5	0.620	0.750
	女	22	2		
年龄	<45 岁	50	5	0.020	0.150
	≥45 岁	14	2		
初诊时 WBC	≥10 × 10 ⁹ /L	18	6	<0.001	0.005
	<10 × 10 ⁹ / L	46	1		
初诊时 BPC	<20 × 10 ⁹ / L	48	5	0.450	0.670
	≥20 × 10 ⁹ /L	16	2		
形态学	细颗粒	39	4	0.520	0.470
	粗颗粒	25	3		
RAS	曾发生	9	1	0.160	0.670
	未发生	55	6		

【案例释疑】 正确的做法应多选择几种筛选自变量的方法建立 logistic 回归模型和 Cox 回归模型，并借助一些评价方法，确定一个最合适的方程，据此得出专业结论。

9.5.5 本节相关知识简要回顾

进行线性相关与回归分析时应注意要依专业知识选择分析指标，拟分析的数据要具有同质性。进行线性相关与回归分析时应画散点图，根据散布图的变化趋势，决定是进行直线相关与回归分析还是进行某种曲线回归分析，如进行直线相关与回归分析应对计算的相关系数和回归系数进行假设检验，通常认为 $R^2 \geq 0.5$ 时相关分析才具有一定的实际意义。

对于常用的三大回归(多重线性回归、多重 logistic 回归和多重 Cox 回归分析)，当协变量个数较多时可进行变量的筛选，常用的筛选变量的方法有前进法、后退法和逐步回归法。在进行三大回归分析时，人们易犯的错误有：将性质截然不同的数据混在一起进行多重回归分析；筛选变量的策略有误，仅把经单因素分析具有统计学意义的变量纳入三大回归方程分析；采取不筛选变量的方法得到回归方程，故结果中仍包含很多无统计学意义的自变量，因而所得结论可信度低。回归方程中的变量及其偏回归系数估计值的正负号应符合专业意义。

9.6 统计结果报告中的常见问题及案例辨析

9.6.1 统计结论写法有误

【案例 9-32】 检测特发性血小板减少性紫癜(ITP)患者及非免疫性血小板减少患者分泌 GP II b/Ⅲa 抗体 B 细胞、血小板特异性抗体的变化，评价其对诊断 ITP 及非免疫性血小板减少疾病的作用及临床意义。结果部分：ELISPOT 法检测分泌 GP II b/Ⅲa 抗体 B 细胞频数用于 ITP 诊断有显著意义($P=0.000$)。

【案例辨析】 第一，原文有显著意义的表达不合适。统计学检验只能回答有无统计学意义，是否有显著意义需要根据专业意义回答；第二， $P=0.000$ 的表达不合适，SPSS 软件给出 $P=0.000$ 的表达是由于软件只给出小数点后三位的值；第三，原文的统计结果只给出 P 值，而没有给出相应的检验统计量的数值。

【案例释疑】 第一，原文统计结果应表达为有统计学意义；第二， $P=0.000$ 应写为 $P<0.001$ ；第三，原文的统计结果不仅需给出 P 值，而且应给出相应的检验统计量的数值。

【案例9-33】 回顾性分析非血缘脐血移植(UCBT)治疗成人恶性血液病患者的植入、移植相关并发症及生存情况。部分结果:多份UCBT植入脐血所含NC与单份脐血移植患者所输NC有明显差异($P=0.005$),两者所含CD34⁺细胞数、CFU数差异不明显。

【案例辨析】 第一,原文中有明显差异的表达不合适,统计学检验只能回答有无统计学意义,是否有明显差异需要根据专业意义回答;第二,原文的统计结果只给出 P 值,而没有给出相应的检验统计量的数值。

【案例释疑】 第一,原文统计结果应表达为有统计学意义;第二,原文的统计结果不仅需给出 P 值,而且应给出相应的检验统计量的数值。

9.6.2 统计表有误, P 值一般写在表外

【案例9-34】 评价再生障碍性贫血(AA)患者骨髓造血细胞遗传不稳定性,探讨其对患者免疫抑制治疗(IST)的影响。选择2005年3月至2008年4月在我院住院治疗的AA患者204例,其中重型AA(SAA)+极重型AA(VSAA)患者144例,治疗前进行彗星试验检测的SAA和VSAA患者中72例接受了IST,随访6个月以上者36例,选取尾矩(TM)、Olive尾矩(OTM)和彗星细胞率(Comet%)作为分析指标,见表9-27。

表9-27 36例SAA和VSAA患者遗传不稳定性与IST近期疗效[例(%)]

组 别	完全反应	部分反应	无 反 应	P 值	组 别	完全反应	部分反应	无 反 应	P 值
Comet% 分组					OTM 分组				
<21.64%	6(31.6)	6(31.6)	7(36.8)	0.923	<1.58	5(26.4)	7(36.8)	7(36.8)	0.923
≥21.64%	3(17.6)	8(47.1)	6(35.3)		≥1.58	4(23.5)	7(41.2)	6(35.3)	

部分结果:Comet% <21.64%组有效率63.2%(19例中12例),Comet% ≥21.64%组有效率64.7%(17例中11例);OTM分组<1.58有效率63.2%(19例中12例),OTM分组≥1.58组有效率64.7%(17例中11例)。

【案例辨析】 表9-27中结果变量疗效是多值有序变量,因而属结果变量为有序变量的单向有序二维列联表资料。原文作者将结果变量疗效分为完全反应、部分反应和无反应三个等级,又将完全反应和部分反应两个水平进行合并,使用 χ^2 检验进行分析,这样做损失了数据的有序性信息,同时需注意一般不在统计表内标注统计分析的结果即检验统计量及对应的 P 值。

【案例释疑】 第一,宜采用秩和检验分析该资料;第二,删掉统计表9-27中的 P 值,在表外用文字描述检验统计量和 P 值(注意:若结果项目少时,可用备注的形式将其写在表底;若结果项目多时,应用一段文字专门描述结果并给出统计和专业结论),此表修改为表9-28;第三,若原作者特别希望从专业角度用“率”来分析数据,可在获得正确分析结果及解释之后,附加上按率分析的结果和解释。

表9-28 36例SAA和VSAA患者遗传不稳定性与IST近期疗效[例(%)]

组 别	类 型	例 (%)		
		疗效: 完全反应	部分 反 应	无 反 应
Comet% 分组	<21.64%	6(31.6)	6(31.6)	7(36.8)
	≥21.64%	3(17.6)	8(47.1)	6(35.3)
OTM 分组	<1.58	5(26.4)	7(36.8)	7(36.8)
	≥1.58	4(23.5)	7(41.2)	6(35.3)

9.6.3 用“均值 ± 标准误差”描述正态分布资料的分布特征

【案例 9-35】 探讨角质细胞生长因子(KGF)在小鼠异基因脐血移植(UCBT)后免疫重建中的作用及其机制。统计学处理：计量资料以均值 ± 标准误差表示。

【案例辨析】 原作者在描述定量指标的结果时采用均数 ± 标准误差的形式。正态分布定量资料用均值 ± 标准差($\bar{x} \pm s$)描述其分布特征，用均值 ± 标准误差($\bar{x} \pm s_x$)不妥当，由于 $s_x = \frac{s}{\sqrt{n}}$ ，即使标准差远大于算术均数(此类资料不服从正态分布，甚至也不呈对称分布)，其标准误差也可能较小。

【案例释疑】 通常情况下，以“均数 ± 标准差”的形式描述呈正态分布定量(至少也应呈对称分布)资料，如果标准差明显大于算术均数时，就采用 $M(Q_1 \sim Q_3)$ (或 $M(QR)$)的形式描述该定量资料。

9.6.4 用“均值 ± 标准差”描述偏态分布资料的分布特征

【案例 9-36】 通过监测急性白血病患者柔红霉素化疗前后血浆前脑利钠肽(NT-pro-BNP)及心肌酶谱(CK、CKMB)水平变化并进行对照研究，以探讨 NT-pro-BNP 对柔红霉素引起患儿心肌损伤的敏感程度的评价作用。统计学处理：计量资料采用 $\bar{x} \pm s$ ，自身对照研究采用配对 t 检验。结果部分见表 9-29。

表 9-29 应用柔红霉素前后血浆(NT-pro-BNP)及 CK、CKMB 比较($\bar{x} \pm s$)

组 别	例 次	NT-pro-BNP (ng/L)	CK (U/L)	CKMB (U/L)
治疗前	79	103.3 ± 89.1	37.6 ± 52.4	17.3 ± 10.1
治疗后	79	186.3 ± 171.6	43.0 ± 54.5	19.0 ± 15.3
t 值		4.395	1.742	1.797

注：经配对资料 t 检验；^a $P < 0.01$ ；^b $P > 0.05$ 。

【案例辨析】 第一，从表 9-29 中可知治疗前后 CK 的标准差都大于均数，故治疗前后的 CK 值可能不服从正态分布，不宜用 $\bar{x} \pm s$ 进行统计描述；第二，要进行配对 t 检验，每一指标的差值 d 需满足正态分布，原文中并未提到对差值 d 进行正态性检验。

【案例释疑】 第一，通常情况下，以“均数 ± 标准差”的形式描述呈正态分布(至少呈对称分布)定量资料，如果标准差明显大于算术均数时，应采用 $M(Q_1 \sim Q_3)$ (或 $M(QR)$)的形式来表达该定量资料；第二，在应用配对设计定量资料 t 检验分析资料时，差值 d 需要满足正态性，差值 d 不满足正态性，可对差值 d 进行某种变量变换，若经变量变换后的资料满足正态性，仍可采用配对设计定量资料 t 检验分析资料，否则，只能选择适合的非参数检验方法，如配对设计符号秩检验。

9.6.5 误用条图表达适合用线图表达的资料——横轴上的变量为时间且横轴刻度表达有误

【案例 9-37】 探讨骨髓间充质干细胞(MSC)对半相合骨髓移植小鼠免疫功能的影响及机制。部分结果见图 9-2 和图 9-3。

【案例辨析】 第一，复式条图适合表达具有一个统计指标，两个分组因素的资料，横轴适合安排其中的一个分组因素，本例横轴所示的移植时间是一个连续型变量，因而不适合用复式条图来表达；第二，图 9-2、图 9-3 中等长的间距并不代表相等的数量。

【案例释疑】 第一，当希望表达一个或多个事物随时间推移的变化规律，宜用普通线图，横轴

表示时间,纵轴表示观察指标;第二,绘制统计图时,对于算术尺度,要求坐标轴上等长的间距代表相等的数量。

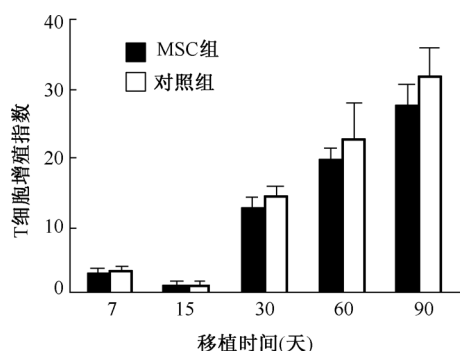


图 9-2 移植后 ConA 刺激受者 T 细胞增殖情况

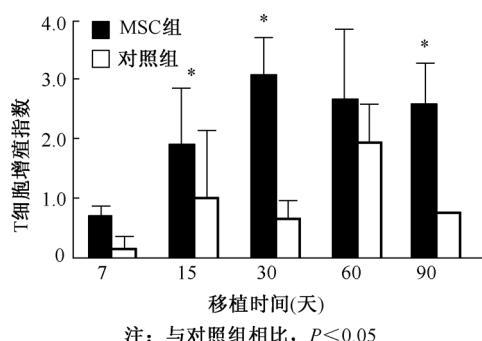


图 9-3 移植后 LPS 刺激受者 B 细胞增殖情况

9.6.6 误用条图表达适合用线图表达的资料——横轴上的变量为浓度

【案例 9-38】 研究苦参碱诱导 K562 细胞分化过程中 IER3IP1 基因的表达式,以明确 IER3IP1 基因表达式与苦参碱作用的量效及时效关系,并初步探讨该基因在 K562 细胞中的功能。部分结果见图 9-4。

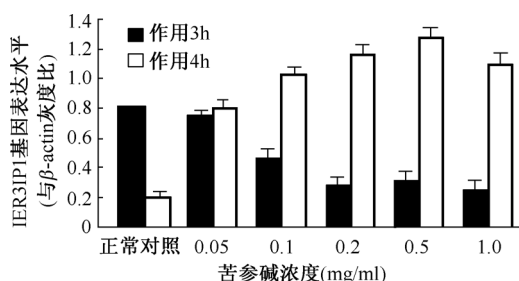


图 9-4 不同浓度的苦参碱作用 K562 细胞 3h 和 24h 后 IER3IP1 基因的表达水平

【案例辨析】 第一,复式条图适合表达具有一个统计指标,两个分组因素的资料,横轴安排其中的一个分组因素,本例横轴所示的苦参碱浓度是一个连续型变量,要反映事物或现象随浓度的变化,不适合用条图表达,因为条图割断了浓度之间的联系;第二,图 9-4 中等长的间隔代表的苦参碱浓度不相等。

【案例释疑】 第一,当希望表达一个或多个事物随浓度的变化规律,宜用普通线图,横轴表示浓度,纵轴表示观察指标;第二,绘制统计图时,对于算术尺度,要求坐标轴上等长的间距代表相等的数量。

9.6.7 本节相关知识简要回顾

描述呈正态分布或对称分布定量资料的分布特征用均数 \pm 标准差,描述呈偏态分布资料的分布特征用 $M(Q_1 \sim Q_3)$ (或 $M(QR)$)。根据资料的性质和分析目的正确选择并绘制统计表与统计图,避免常犯错误,如用条图表达适合用线图表达的资料——横轴为时间或浓度。统计结果中尽可能给出检验统计量的数值和相应的 P 值,统计结论为有无统计学意义,而非有无明显差异,是否有明显差异需要根据专业知识回答。

9.7 本章内容小结

正确应用医学统计方法直接关系到医学科研论文的质量,然而,在医学期刊中可以见到多种统计学应用方面的错误。本章的原始资料摘录于《中华血液学杂志》,本章不仅对发表于该杂志的文章所用到的统计分析方法、统计软件和研究内容进行归类总结,而且重点对文章所涉及的科研设计、统计分析和结果报告中的常见错误进行辨析与释疑。本章通过分析发表于该杂志的文章中的统计学错误,促使科研工作者重视科研设计与统计分析知识,进而提高统计应用水平,加强科研工作者在论文撰写方面的科学性和严谨性。

参 考 文 献

- [1]胡良平. 科研课题的研究设计与统计分析——错误案例辨析与释疑. 北京: 军事医学科学出版社, 2008: 43-420.
- [2]胡良平. 科研课题的研究设计与统计分析——提高学位论文统计学质量的对策. 北京: 军事医学科学出版社, 2010: 91-348.
- [3]胡良平. 统计学三型理论——在统计表达与描述中的应用. 北京: 人民军医出版社, 2008: 65-317.
- [4]陆再英, 钟南山. 内科学. 北京: 人民卫生出版社, 2009: 564-565.
- [5]刘桂芬. 医学统计学. 北京: 中国协和医科大学出版社, 2010: 41-269.
- [6]孙振球. 医学统计学. 北京: 人民卫生出版社, 2006: 235-258.

第 10 章 连续型随机变量概率分布与基因数据分析

对于应用者而言,了解统计学中的常用分布及其内在的关系是很有必要的。因此,本章首先对卫生统计学中常遇到的基本统计分布进行简单介绍,包括正态分布、Student's t 分布、 χ^2 分布和 F 分布。其次以总体均数估计和单组设计一元定量资料 t 检验为例详细阐述统计学中几个最基本的概念:点估计、区间估计和 P 值。一些新兴医学技术的发展和應用导致了实际工作中更容易得到高通量(High Throughput)的数据,如微阵列技术(Microarray Technology),高通量数据对传统统计学方法提出了前所未有的挑战,很多新的统计分析方法被提出来解决高维数据分析问题。再次介绍微阵列数据中关于基因差别表达分析的多重假设检验(Multiple Hypothesis Testing)问题;最后是相应的实际例子和具体分析及结果的表达和解释。

10.1 常用统计分布

10.1.1 正态分布

在所有的统计分布中,正态分布(Normal Distribution)是最为常见的,因为它是许多其他分布的极限分布。设 x 是一个一元随机变量,若 x 的概率密度函数(Probability Density Function, PDF)可表示为:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] \quad (10-1)$$

简记作:

$$x \sim N(\mu, \sigma^2)$$

概率密度函数是一个不易被非数学工作者理解的概念,它可以被视为给定 $x = x_0$ 时曲线的纵坐标值 $f(x_0)$,它似乎与“概率密度”无任何关联。事实上,当人们希望知道 x 在 $(x_0, x_0 + \Delta x)$ 这个小区间上取值的概率是多少时,只要这个小区间长度 Δx 充分地小,其取值的概率就可近似地等于 $\Delta x \cdot f(x_0)$ 。显然,固定小区间的长度,纵坐标值越大,其概率值也就越大。由此可以形象地将纵坐标值的大小比作落入指定区域内的雨水的多少,雨水越密集,单位时间内单位区域上的雨量就越多。

在式(10-1)中, μ 和 σ^2 分别是总体均数和方差,分别被称为位置参数(Location Parameter)和尺度参数(Scale Parameter)。位置参数 μ 反映了随机变量在横轴的中心位置,尺度参数 σ^2 刻画了随机变量分布的形状。对于一个具体的样本而言,总体参数 μ 和 σ^2 常常是未知的,需要根据样本来估计,具体内容见下面相关介绍。

特殊地,若 $\mu = 0$, $\sigma^2 = 1$,则 $f(z)$ 被称为标准正态分布(Standard Normal Distribution),记作 $z \sim N(0, 1)$ 。图 10-1 给出了正态分布的概率密度函数图。

正态分布的重要作用主要在于:①正态分布是其他许多分布的基础;②中心极限定理表明,在满足一般正规条件下,大多数统计量的极限分布都是正态分布。医学中常见的指标如身高、体重、红细胞数和脉搏数等皆近似服从正态分布。

在图 10-1 中,与 $N(0, 1)$ 对应的那条标准正态概率密度曲线的使用价值最大,通过它可以方便地计算出随机变量落入横坐标轴上任何区间内的概率。例如, x 落入区间 $[-2.576, 2.576]$ 、

$[-1.960, 1.960]$ 和 $[-1, 1]$ 的概率分别约为 0.990、0.950 和 0.683。换句话说, 任何一个实际问题, 只要能转化成标准正态分布, 就可灵活运用此概率密度曲线得到很多有价值的结果和结论。

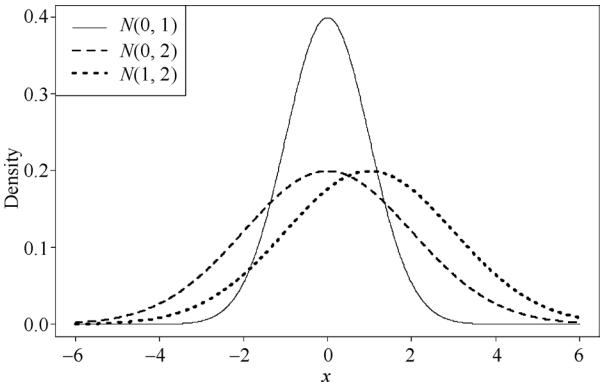


图 10-1 正态分布 $N(0,1)$ 、 $N(0,2)$ 、 $N(1,2)$ 的概率密度函数图

10.1.2 χ^2 分布

另一个常见的分布是 χ^2 分布 (Chi-Square Distribution), χ^2 分布可由正态分布定义。 ν 个标准正态随机变量 z 的平方和, $x = \sum_{i=1}^{\nu} z_i^2$ 服从 χ^2 分布, 记作:

$$x \sim \chi^2_{\nu}$$

χ^2 分布只有一个参数即自由度 ν , 为了纪念统计学家 Pearson (1857 年 3 月 27 日 – 1936 年 4 月 27 日), 又称之为 Pearson χ^2 。 χ^2 分布在率和构成比等定性资料分析和拟合优度检验中起着重要作用。 χ^2 分布的概率密度函数见图 10-2, 随着自由度的增大, χ^2 分布逐渐趋向对称。

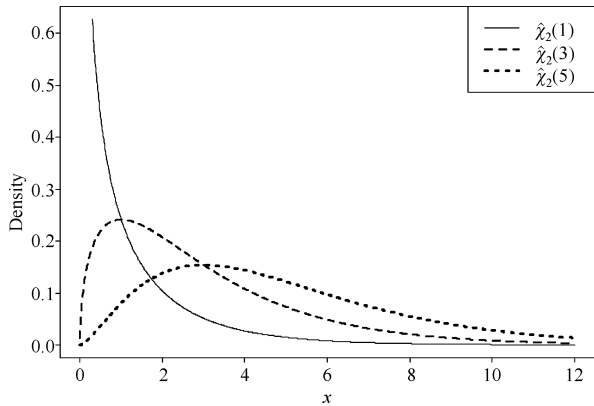


图 10-2 χ^2 分布 $\nu=1, 3, 5$ 的概率密度函数图

10.1.3 t 分布

t 分布由 W. S. Gosset (1876 年 6 月 13 日 – 1937 年 10 月 16 日) 在 1908 年以笔名 “Student” 发表, 即著名的 Student’s t 分布。 t 分布的最初叫法来源于 Fisher 1924 年的文章, Fisher (1890 年 2 月 17 日 – 1962 年 7 月 29 日) 于 1923 年给出了 t 分布的推导, 可由标准正态分布和 χ^2 分布定义得到。统计学史上, 以此为界将统计学发展分为描述统计学和推断统计学, 前者的代表是 Pearson, 后者的代表是

Fisher 和 Neyman(1894 年 4 月 16 日 - 1981 年 8 月 5 日) 等人。设 z 为标准正态随机变量, x 服从自由度为 ν 的 χ^2 分布: $x \sim \chi^2_\nu$, 并且两者相互独立, 则 t 分布定义为:

$$t = \frac{z}{\sqrt{x/\nu}} \sim t(\nu) \quad (10-2)$$

t 分布只有一个参数即自由度 ν , 记作 $t \sim t(\nu)$ 。 t 分布的重要意义在于使得小样本的统计推断变成可能, 统计学发展进入一个更加精细和全新的时期。 t 分布的概率密度函数见图 10-3, 随着 ν 的增大, t 分布越来越接近标准正态分布, $\nu = \infty$ 时的 t 分布就是标准正态分布。

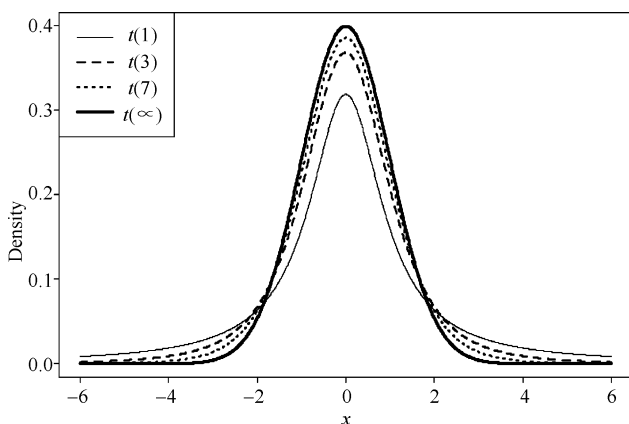


图 10-3 t 分布 $\nu=1, 3, 7$ 和 ∞ 的概率密度函数图

作为一个简单的例子, 下面介绍单组设计一元定量资料 t 检验的公式:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{\nu=n-1} \quad (10-3)$$

此公式的来历如下。样本 $X = (x_1, \dots, x_n)$ 来源于正态总体分布 $N(\mu, \sigma^2)$, 则:

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$\bar{x} = \sum_{i=1}^n x_i/n$ 为样本均数, n 为样本量。由于总体标准差 σ 未知, 用样本标准差 s 估计总体标准差 σ , 由于 s 取代了 σ 的位置, 有限样本下 $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ 的分布不再是标准正态分布了。由简单数理统计知识得知:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{\nu=n-1} \quad (10-4)$$

再根据 t 分布的定义[见式(10-2)], 得到常见的单组设计一元定量资料 t 统计量公式:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2/\sigma^2}{(n-1)}}} \quad (10-5)$$

式(10-5)的分子为标准正态分布, 分母为一个 χ^2 分布和自由度比值的平方根。由式(10-2)及其有关的定义可知, 由式(10-5)所定义的随机变量 t (同时它也被称为统计量)服从自由度为 $n-1$ 的 t 分布。

10.1.4 F 分布

F 分布是方差分析的理论基础, 为了纪念 Fisher, 故称为 F 分布, F 分布的名称由 Snedecor (1881 年 10 月 20 日 - 1974 年 2 月 15 日) 于 1932 年引入, 可以由 χ^2 分布推导得到。设两个独立的 χ^2 变量: $y_1 \sim \chi_n^2, y_2 \sim \chi_m^2$, F 分布定义为:

$$F = \frac{y_1/n}{y_2/m} \sim F(n, m) \quad (10-6)$$

式中, n, m 分别是 F 分布的分子 (第一) 自由度和分母 (第二) 自由度, 记作 $F \sim F(n, m)$ 。特殊地, 自由度为 n 的 t 分布等价于第一自由度为 1、第二自由度为 n 的 F 分布。 F 分布的概率密度函数见图 10-4, 随着第一和第二自由度的增大, F 分布越来越接近对称分布。

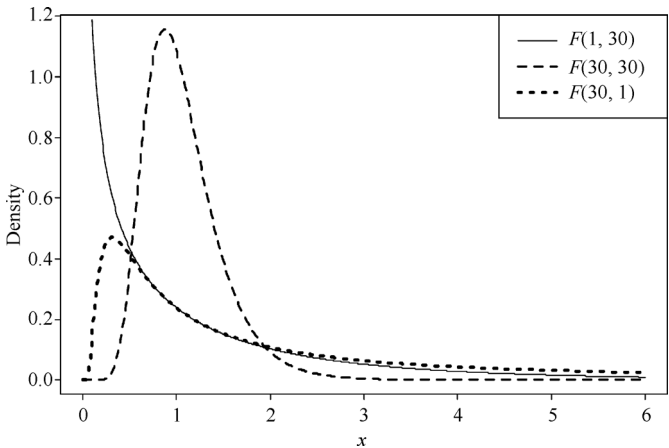


图 10-4 F 分布的概率密度函数图

10.2 置信区间和 P 值

置信区间和 P 值是统计学中两个典型的具有哲学意义的、极富争议的概念, 本节主要对这两个概念进行详细解释, 这种解释有利于读者理解其他章节的内容。

10.2.1 基本概念

统计分析包括统计描述和统计推断, 关于统计描述的内容在此不作介绍, 统计推断包括参数估计 (Parameter Estimation) 与假设检验 (Hypothesis Testing), 其中参数估计又包括点估计 (Point Estimation) 和区间估计 (Interval Estimation)。点估计是指用随机样本的样本统计量的值直接作为相应总体参数估计值的方法, 即直接用样本统计量的值代替相应总体参数值, 不考虑抽样误差的影响。区间估计是指用已知样本统计量和标准误差确定一个有概率意义的区间, 且该区间具有较大可信度 (置信度) $1 - \alpha$ 包含总体参数的估计方法, 该区间称为总体参数的置信区间或可信区间 (Confidence Interval, CI)。置信度 $1 - \alpha$ 常取为 95%, 若是双侧 95% 置信区间, 即区间位于某特定概率密度曲线之下且在横坐标上居中的那段范围, 而随机变量在此范围内取值的概率为 95%。

10.2.2 区间估计

这里只考虑单个参数的区间估计问题, 如总体均数 μ 。正态总体分布包括位置参数 μ 和尺度参

数 σ^2 , 当估计总体均数 μ 时, σ^2 是不能不考虑的。一般而言 σ 是未知的, 用样本标准差 s 作为 σ 的估计值计算标准误 $s_{\bar{x}}$, 按 t 分布原理, 有:

$$P\left(-t_{1-\alpha/2, \nu} < \frac{\bar{x} - \mu}{s_{\bar{x}}} < t_{1-\alpha/2, \nu}\right) = 1 - \alpha, \nu = n - 1 \quad (10-7)$$

$$\bar{x} - t_{1-\alpha/2, \nu} s_{\bar{x}} < \mu < \bar{x} + t_{1-\alpha/2, \nu} s_{\bar{x}}$$

式中, $1 - \alpha$ 表示置信度, $t_{1-\alpha/2, \nu}$ 表示自由度为 ν 、双侧尾部面积为 α 的 t 分布界值。常取 $\alpha = 0.05$ 。表 10-1 给出了 10 个 TB/HIV 双重感染者的 CD₄ 细胞计数值 (经对数转换后), 详细地说明见后文。

表 10-1 10 个 TB/HIV 双重感染者 CD₄ 细胞计数的对数转换值

0.69	2.08	5.35	4.61	6.02	5.80	2.20	4.29	3.04	2.56
------	------	------	------	------	------	------	------	------	------

如果假设这 10 个 TB/HIV 双重感染者的 CD₄ 细胞计数对数值是从正态总体分布中随机抽样得到的 (验证这一假设需要另外的工作, 关于如何检验一个总体是否服从某一个理论分布可参见其他章节或相关的统计文献), 根据式 (10-7) 可计算 CD₄ 对数值总体均数 μ 的置信区间:

$$\begin{aligned} \bar{x} &= 3.66, s_{\bar{x}} = 0.57, t_{1-\alpha/2, \nu} = t_{1-0.05/2, 9} = 2.62 \\ \bar{x} - t_{1-\alpha/2, \nu} s_{\bar{x}} &= 3.66 - 2.62 \times 0.57 = 2.37 \\ \bar{x} + t_{1-\alpha/2, \nu} s_{\bar{x}} &= 3.66 + 2.62 \times 0.57 = 4.56 \end{aligned} \quad (10-8)$$

因此, 可认为 TB/HIV 双重感染者的 CD₄ 细胞计数对数值的总体均数 95% 置信区间为 [2.37, 4.56]。

10.2.3 P 值

如果进一步想知道这些 TB/HIV 双重感染者的 CD₄ 细胞计数对数值的平均值是否与另一个人群的 CD₄ 细胞计数对数值的平均值有差别, 同样在正态分布条件下, 并假设另一个人群的 CD₄ 细胞计数对数值的均值为 $\mu_0 = 5.58$ 。这里需要说明的是, 上述 TB/HIV 双重感染者和这里的“另一个人群”实际上是从一个样本量 $n = 315$ 的资料中得到的, 该资料来源于山西省运城市第五轮中国全球基金结核病项目芮城、夏县、新绛、绛县、稷山 5 个项目防治县 TB/HIV 双重感染者进行干预与治疗的监测数据, 其中 TB/HIV 双重感染者 10 人, 无 TB 感染者 305 人, 由于该组人群样本量大, 所以直接应用无 TB 感染者 CD₄ 细胞计数对数值的均值作为对照组均值而不考虑抽样误差。

问题的实质是, 根据样本的信息 (10 个 TB/HIV 双重感染者的 CD₄ 细胞计数对数值的均值, 见表 10-1) 推断该样本所代表总体——假设均值为 μ , 与另外一个已知总体的均值 μ_0 是否有差别, 只需要进行一个单组设计一元定量资料 t 检验:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3.66 - 5.58}{0.57} = -3.36$$

对应的双侧 P 值为:

$$P = \text{Prob}(|T| \geq t \mid H_0: \mu = \mu_0) \approx 0.008$$

P 值可能是所有实际应用者最关心的问题。通常的法则, 是将 P 值和检验水准 α (一般取 0.05) 作对比, 如 $P \leq \alpha$, 则做出拒绝 H_0 的决定, 否则没有理由拒绝 H_0 。显然, 此处做出拒绝 H_0 的决定。

在上述分析过程中, 构成检验假设的两个对立方面分别是:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ H_1: \mu &\neq \mu_0 \end{aligned}$$

而 P 值究竟是什么呢? 其意义在于: 若将上述调查重复一千次、一万次, 甚至更多次, 每次调查都

能得到一个对应的检验统计量的 t 值(上面一次调查结果中 $t = 3.36$), 在所获得的如此多的 t 值中, 其绝对值大于等于 3.36 的概率 P 为 0.008。

10.2.4 贝叶斯统计和频率统计

一个重要的问题是, 上述的置信区间[2.37, 4.56]和 $P = 0.008$ 该如何解释? 包括麻醉学期刊在内的许多医学期刊论文都存在理解上的错误, 常常出现如下的错误解释, 例如:

置信区间 [2.37, 4.56]: 认为 TB/HIV 双重感染者 CD_4 对数值的总体均数 μ 在此区间的概率为 95%。

$P \approx 0.008$: H_0 正确的可能性不到 0.8%, 因此拒绝 H_0 , 进而, 由于 P 值太小, 因此认为有高度显著的意义。

例如, “烫伤后 6h, 与 DS 组比较, DB 组、SUF 组和 NAL 组血浆 MDA 和 cTnI 的含量显著升高 ($P < 0.01$), SOD 活性显著降低 ($P < 0.01$); 与 SUF 组比较, DB 组、NAL 组血浆 MDA 和 cTnI 的含量显著升高 ($P < 0.01$), SOD 活性显著降低 ($P < 0.01$)”, 同样的陈述还常见于其他文献。

然而, 要彻底地回答这一问题并不简单, 需要涉及统计学中最基本的哲学问题: 参数是否是随机的?

对这一问题的回答, 统计学中存在两种截然不同的看法: 频率统计(Frequency Statistics), 也即是大多数教科书所讲授的知识, 认为参数是客观存在且固定不变的, 不存在随机性的问题; 贝叶斯统计(Bayesian Statistics)——国内的教科书中基本没有介绍, 和经典频率统计的最大不同之处在于, 认为参数是随机变量而非固定的未知常数, 并且认为概率是对随机事件的主观信任度(Degrees of Belief), 也即是概率表示人们认为事件为真的程度。假设希望从数据中估计参数 θ , 这里参数可以是一切未知量, 如回归系数、均数或缺失值(Missing Data)等, 相对频率统计而言, 贝叶斯统计认为在收集到数据之前, 或多或少地具有参数 θ 的已知信息, 并且可以用概率分布反映拥有信息的不确定性, 这个分布称为先验分布(Prior Distribution)。在得到数据后根据数据来改变和更新关于参数 θ 的认识, 建立在先验和似然之上的分布称为 θ 的后验分布(Posterior Distribution), 后验分布包含了当前阶段关于参数所有的信息, 是贝叶斯统计推断的出发点。

大体而言, 贝叶斯分析可以分为三个过程: ① 通过先验对参数已有信息做出概括(包括有先验信息和无先验信息)和用似然模型描述收集到资料的数据产生机制。② 根据贝叶斯定理得到后验分布, 并计算后验统计量和推断或预测。在复杂的贝叶斯模型中, 这一步往往通过模拟尤其是应用马尔科夫蒙特卡洛(Markov Chain Monte Carlo, MCMC)算法实现。③ 解释后验分布的实质意义和进行拟合优度评价, 也即是核查所选择模型的假设是否与实际观察数据相符, 在明显偏离数据时需要重新选择模型(或/和先验)并重复上述三个过程。

10.2.5 置信区间和 P 值的频率统计解释

显然, 因为认为参数是随机的, 从贝叶斯统计的角度看, 上述的解释是合理的。但是, 如何正确理解置信区间和 P 值的频率统计学意义呢? 所有频率统计学基本上都置于重复抽样的框架下进行解释。

置信区间的具体含义并非针对某一个根据样本具体计算出的区间而言, 如[2.37, 4.56], 而是针对计算这个置信区间的公式, 或者更准确地说, 是针对某种具体的统计学方法而言的, 也即是需要对式(10-7)做出解释, 具体含义:

如果重复地从与这 10 个 TB/HIV 双重感染者相同的总体中抽样, 每次抽取 10 个, 按照式(10-7)计算总体均数 μ 的置信区间, 那么, 平均而言, 将有 95% 的区间会包含总体均数 μ 。

图 10-5 展示了这一过程。从 $N(3.66, 1.81)$ 的正态分布中重复抽样, 每次 10 例, 按照式 (10-7) 计算总体均数 μ 的置信区间, 其中约有 5 个置信区间不包含真实参数 3.66, 见图中三角符号所示之处。

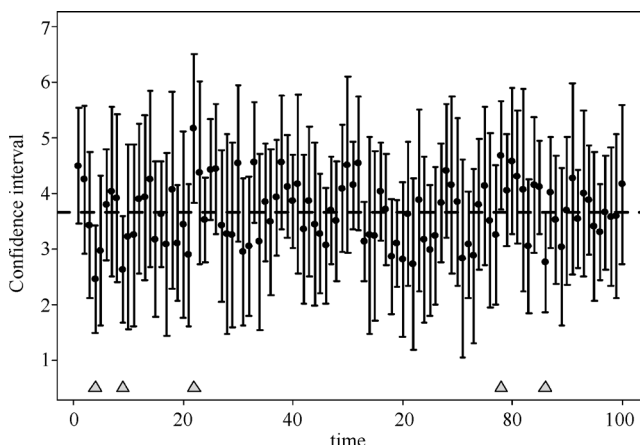


图 10-5 TB/HIV 双重感染者 CD_4 对数值的 95% 置信区间示意图

相应地, P 值的含义是:

重复地从与这 10 个 TB/HIV 双重感染者相同的总体中抽样(这暗示, 双重感染者和对照具有相同的参数), 每次抽取 10 个, 按照式 (10-4) 执行单组设计一元定量资料 t 检验, 这样就会得到很多性质相同的 t 值, 记为 t^i , P 值的计算为:

$$P = \lim_{N \rightarrow \infty} \frac{\#\{i: |t^i| \geq 3.36; i = 1, 2, \dots, N\}}{N}$$

即, 绝对值大于等于现有 t 统计量(绝对值)的比例。

图 10-5 展示了这一过程。假设 TB/HIV 双重感染者 CD_4 细胞计数对数值的总体均数 $\mu_0 = \mu = 5.58$, 总体标准差 $\sigma_0 = s = 1.81$ 。从总体 $N(5.58, 1.81)$ 中抽取含量为 10 的样本, 按照式 (10-4) 执行单组设计一元定量资料 t 检验, 重复 10^6 次, 其中绝对值大于或等于 3.36 的比例为 8.43%。与直接根据式 (10-4) 计算的 P 值很接近。总体设为 $N(5.58, 1.81)$ 的原因在于, 假设“ H_0 : TB/HIV 双重感染者 CD_4 细胞计数对数值的平均水平与对照一样”成立, 总体标准差用样本标准差估计。上述过程实际上是一个参数的 Bootstrap 抽样过程。

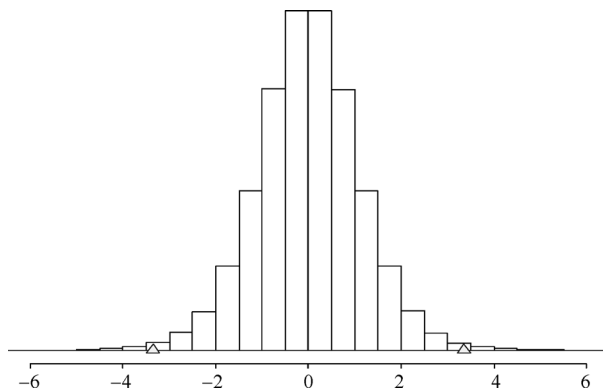


图 10-6 P 值的频率统计解释, 三角符号位置为 ± 3.36

因此, P 值的大小不能作为不同总体间差别程度的度量。仅仅说明了, 在 H_0 成立的条件下, 重复相同的统计检验, 得到现有样本统计量的数值和更极端取值的可能性大小的度量。

10.3 微阵列差别表达分析和多重假设检验

现代生物和医学技术的发展使得人们能够比以往任何时候都能更快地收集到大量的数据, 微阵列 (Microarray) 技术是其中的著名代表, 给科学研究带来了革命性的影响, 为人们提供了一种从没有过的医学实践方式。用含有成千上万种的 DNA 或者蛋白质序列的微小玻璃芯片取代传统生物学中的凝胶、滤器和纯化柱, 微阵列技术使得科学家们能够在基因组规模上对基因表达水平进行快速和定量的检测, 由此产生的大规模数据也给统计科学带来了前所未有的机遇和挑战。图 10-7 展示了一个典型的微阵列过程。

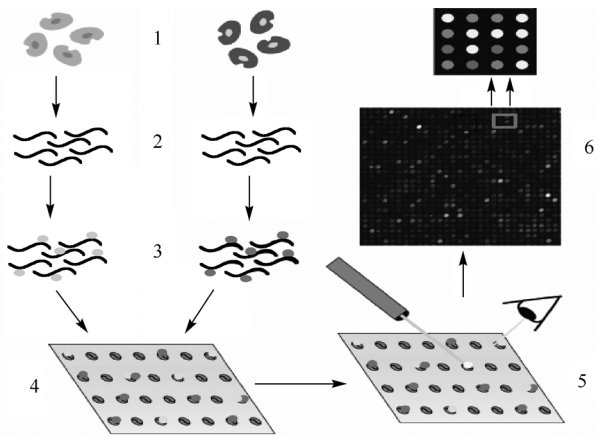


图 10-7 微阵列技术示意图

一个完整的微阵列试验包括以下过程: (1) 选择生物样本 (如正常组织细胞和肿瘤细胞); (2) 提取 mRNA 和逆转录生成 cDNA; (3) 对 cDNA 进行荧光标记制作探针, 通常试验标为红色, 对照标为绿色; (4) 标记的 cDNA 和芯片上的 DNA 序列发生特异性杂交反应, 使得芯片上的点出现荧光信号, 荧光信号的强度和基因表达丰度成正相关; (5-6) 通过激光技术将荧光信号转变为数字信号, 经过前期处理后得到微阵列数据。该图来源于网络。

在微阵列试验中研究者需要同时检验数以千计的基因表达水平是否与不同条件状况之间存在关联, 由此很自然地就涉及了多重假设检验的问题。众所周知, 假设检验中会不可避免地发生 I 型错误或 II 型错误。总体错误率 (Family Wise Error Rate, FWER) 是传统多重假设检验中常用的错误控制指标, 定义为至少犯一次 I 型错误的概率, 但是总体错误率控制过于严格并不适于微阵列数据分析。由 Benjamini 和 Hochberg (1995) 提出的错误发现率 (False Discovery Rate, FDR) 很好地适应了高维数据多重检验的错误控制, 越来越多地应用于微阵列数据分析。本节内容将详细讨论错误发现率的有关问题。

10.3.1 P 值和假阳性率

单个假设检验中主要依靠 P 值 (或者统计量) 做出是否拒绝零假设 H_0 的决定: P 值和预先给定的检验水准 α 作对比, 如果 $P \leq \alpha$ 拒绝 H_0 , 否则不拒绝 H_0 。拒绝 H_0 时主要关心的是 I 型错误, 即假阳性率 (False Positive Rate, FPR), 检验水准 α 给出了事先可能犯 I 型错误的最大概率, 然后在

保证 I 型错误小于或等于 α 的前提下寻找具有最大检验功效的假设检验方法, 后者定义为 $1 - \beta$, β 被称为犯 II 型错误的概率。设在 H_0 条件下, 统计量为 T , t 为对应的观察值, 则 P 值定义为在 H_0 成立的条件下, 重复进行当前的试验, 获得现有统计量及其更极端情况的概率, 表示为:

$$\begin{aligned} P(t) &= \inf_{\Gamma_\alpha: t \in \Gamma_\alpha} \text{Prob}(T \in \Gamma_\alpha \mid H_0) \\ &= \text{Prob}(|T| \geq t \mid H_0) \end{aligned} \quad (10-9)$$

式中, Γ_α 表示由 α 定义的拒绝域, 即 $\text{Prob}(T \in \Gamma_\alpha \mid H_0) = \alpha$, 如前文。当 $T \in \Gamma_\alpha$ 时必定有 $P \leq \alpha$, 因此基于 P 值的结果等价于基于 t 值的结果。 P 值量化了数据所提供的拒绝 H_0 的证据强度, 是拒绝 H_0 时所犯 I 型错误的最小概率。很显然, P 值越小研究者拒绝 H_0 的信心越大。

10.3.2 多重假设检验和总体错误率

当进行多重假设检验时, 每个单独的假设都具有其本身的 I 型错误, 在这种情况下, 如果不进行任何的控制措施, 犯 I 型错误的概率会随着假设检验的个数而迅速增加。例如, 进行两次独立的检验, 犯 I 型错误的概率为 $1 - (1 - \alpha)^2$, 进行 m 次多重检验犯 I 型错误的概率为 $1 - (1 - \alpha)^m$, 随着检验次数 m 的增加犯 I 型错误的可能性将远远大于名义概率 α , 见图 10-8。

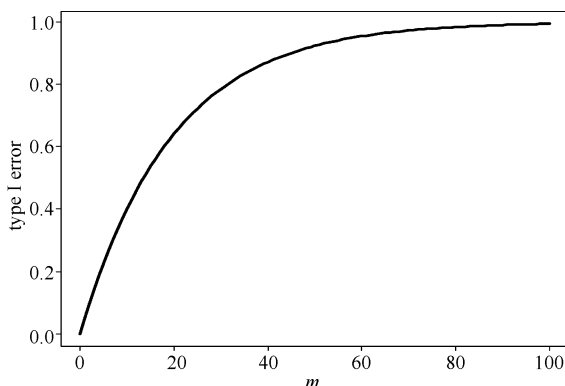


图 10-8 多重比较次数 m 和犯 I 型错误概率的关系

设 m 次多重检验的零假设为 $H_{0i}, i = 1, \dots, m$, 对应统计量分别为 t_i, P_i , 表 10-2 给出了可能出现的检验结果。 m_0 和 m_1 分别表示在 m 次多重检验中真实 H_0 和非真实 H_0 的个数, V 表示在所有 R 次拒绝 H_0 的决定中拒绝了原本真实 H_0 的次数。除 m, R 和 W 外, 表 10-2 中的其他量均为没有观察到的随机变量。在微阵列差别表达分析中, m 即是基因数, m_0 和 m_1 分别为无差别和有差别表达的基因数。

多重检验中广泛使用的错误率控制指标是总体错误率 (Family-Wise Error Rate, FWER):

$$\text{FWER} = \text{Prob}(V \geq 1) \quad (10-10)$$

也即是至少出现一次错误地拒绝真实 H_0 的可能性。各种不同的控制程序用来保证在多重检验时 $\text{FWER} \leq \alpha$, 如 Bonferroni 法: 当 $P_i \leq \alpha/m$ 时拒绝 H_{0i} 。其他改进方法包括 Holm、Hommel、Hochberg 程序。

表 10-2 m 次多重检验的可能结果

	频 数		
	不拒绝 H_0	拒绝 H_0	合 计
H_0 为真	U	V	m_0
H_1 为真	T	S	m_1
合计	W	R	m

对微阵列数据的差别表达分析而言总体错误率控制并不合适, 原因主要如下。

(1) 在经典的多重检验中, m 通常较小, 控制总体错误率具有实质性的意义, 但在微阵列数据中总体错误率控制过于严格, 例如对于 Bonferroni 法, 在 m 很大的情况下 (微阵列数据中 m 常常以

千计,如 Singh et al. 2002 的前列腺癌微阵列数据包含了 50 个正常人和 52 个癌患者 6033 个基因的表达数据,见后文),只有当 P 值很小才能获得一个具有差别表达的基因。

(2)微阵列分析多作为一种探索性的分析策略,主要目的是为后续研究提供候选基因, m 一般介于几百到几万之间,此时控制总体错误率显然没有多大意义。

因此相对于控制总体错误率,研究者更关心的是能否尽量多地识别出差异表达的基因,并且能够容忍和允许在 R 次拒绝中发生少量的错误识别,称为错误发现(False Discovery)。这些原因决定了在微阵列数据中需要在错误发现 V 和总的拒绝次数 R 之间寻找一种平衡,即在检验出尽可能多的候选基因的同时将错误发现控制在一个可以接受的范围内。

10.3.3 错误发现率

在微阵列分析中,相对犯一次或多次 I 型错误的概率,更有意义的是在所有 R 次拒绝中出现错误拒绝 V 的比例。错误发现率(False Discovery Rate, FDR)为上述的平衡提供了一种可能,即 V/R 。FDR 定义为:

$$\text{FDR} = E[V/(R \vee 1)] = E(V/R | R > 0)P(R > 0) \quad (10-11)$$

$R \vee 1$ 表示当 $R=0$ 时 $\text{FDR}=0$, 错误发现率表示了在所有 R 次拒绝中错误发现的期望比例。错误发现率和假阳性率之间有着本质的区别:例如,如果前列腺癌微阵列数据中所有 6033 个基因的表达水平原本是没有差异的(H_0 为真),5% 的假阳性率表示将大约有 $302 = 5\% \times 6033$ 的基因会被错误地认为差别表达;5% 的错误发现率则表示,如果在所有 6033 个多重假设检验中有 100 次拒绝 H_0 (即 $R=100$),那么大约有 $5 = 5\% \times 100$ 基因是被错误识别的。

设预先的错误发现率控制水平为 α (这里的 α 虽然和常用的检验水准 α 具有相同的符号,但是却具有不同的含义),Benjamini 和 Hochberg (BH) 给出了一个基于 p 值的逐步向下控制程序:

- (1) 将 p 值排序: $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$, $H_{0(1)}, H_{0(2)}, \dots, H_{0(m)}$ 为对应的原假设;
- (2) 寻找最大的 k 满足 $p_{(k)} \leq k\alpha/m$;
- (3) 拒绝所有 $H_{0(1)}, H_{0(2)}, \dots, H_{0(k-1)}, H_{0(k)}$, 否则不拒绝任何一个 H_0 。

Benjamini 和 Hochberg (1995) 证明在 BH 控制下, $\text{FDR} = \alpha m_0/m = \pi_0 \alpha \leq \alpha$ 。FDR 允许一定程度的错误拒绝,因此在假设检验次数很多时,具有比 FWER 控制更高的效能。

10.3.4 贝叶斯错误发现率

无疑从 Benjamini 和 Hochberg 的定义看,错误发现率是 V/R 的期望,此外还可以从贝叶斯统计的角度解释错误发现率。设 Γ 为拒绝域,如 $\Gamma = \{Z \leq z\}$,在 $R > 0$ 的条件下[当 m 很大时, $\Pr(R > 0) \approx 1$], FDR 可重新表示为:

$$\begin{aligned} \text{FDR}(\Gamma) &= \Pr(H_0 | Z \in \Gamma) \\ &= \pi_0 F_0(z)/F(z) \end{aligned} \quad (10-12)$$

如 H_0 成立则 $F_0(z)$ 为标准正态变量分布函数, $F(z)$ 为 z 值的分布函数。Efron et al. (2001) 和 Efron (2005) 发展了另一个基于单个统计量的错误发现率,称为局部错误发现率(local false discovery rate, fdr):

$$\text{fdr}(z) = \pi_0 f_0(z)/f(z) \quad (10-13)$$

$f_0(z) = N(0,1)$, $f(z)$ 为 z 值的边际密度。经典的 p 值定义为:

$$p = \inf_{\Gamma: z \in \Gamma} \Pr(Z \in \Gamma | H_0) \quad (10-14)$$

p 值量化了数据所提供的拒绝 H_0 的证据强度,是拒绝 H_0 时所犯 I 型错误的最小概率。FDR(Γ) 量化了 $Z \in \Gamma$ 时拒绝 H_0 所犯的最小 FDR。 p 值的定义基于 H_0 的条件而量化 $Z \in \Gamma$ 的概率, FDR(Γ) 不

过是 p 值的一个逆过程, 基于 $Z \in \Gamma$ 的条件而量化 H_0 成立的概率。从贝叶斯统计的观点看, $\text{FDR}(\Gamma)$ 反应了 $Z \in \Gamma$ 时 H_0 成立的条件概率, 可认为 $\text{FDR}(\Gamma)$ 是贝叶斯后验 p 值 (Bayesian Posterior p Value)。经验贝叶斯方法可用来估计 $\text{FDR}(\Gamma)$:

$$\widehat{\text{FDR}}(z) = \pi_0 F_0(z) / \hat{F}(z) \quad (10-15)$$

$\hat{F}(z) = \#\{Z \leq z\}/m$ 为分布函数的经验估计。Efron 和 Tibshirani (2002) 证明了经验估计和 BH 程序的等价性, 这意味着错误发现率的控制和估计是等价的。定义:

$$z_\alpha = \max(|z| : \widehat{\text{FDR}}(|z|) \leq \alpha) \quad (10-16)$$

那么, 按照 $|z_i| \geq z_\alpha$ 的原则拒绝 H_{0i} 则错误发现率不会超过 α 。

10.3.5 FDR 和总体错误率的关系

能够证明, 在弱控制意义下总体错误率和错误发现率相等, 在强控制意义下总体错误率大于错误发现率。这意味着, 当所有假设检验都为真时, 控制总体错误率也即控制错误发现率, 当 H_0 和 H_1 任意组合时, 控制错误发现率就意味着控制总体错误率。总体错误率控制至少一次错误发现的概率, 从表 10-2 的行“ $m_0 \rightarrow V$ ”考虑问题; 错误发现率控制错误发现比例的期望, 从表 10-2 的列“ $R \rightarrow V$ ”方向出发, 是两个不同的观点和思维。这种看似简单的转换具有突破性的意义, 给大规模数据的多重比较提供了十分恰当的误差测量指标。例如, 控制总体错误率的 Hochberg 程序和 BH 程序具有以下关系:

$$\text{BH/Hochberg} = i/[1 - (i-1)/m] \quad (10-17)$$

由于在大规模数据中研究者可能只对一小部分检验感兴趣, 即 $(i-1)/m$ 很小, 因此式 (10-17) 意味着 BH 程序的标准比 Hochberg 程序宽松约 i 倍。例如, 在 Singh et al. (2002) 的前列腺癌数据中, 按照 0.05 的标准, BH 程序得到 21 个基因, 而 Hochberg 程序只得到 3 个基因。

10.3.6 对相关的敏感程度

大规模数据如微阵列常常存在相关性, 一个自然的问题是, 不同控制程序和估计方法对相关的依赖程度如何? 1995 年的 BH 程序是要求统计量独立的, 但是 Benjamini 和 Yekutieli (2001) 的研究表明, BH 程序在统计量满足所谓正回归相关 (Positive Regression Dependence) 的条件下仍然是有效的, 为了使得 BH 程序在任何相关情况下都有效, Benjamini 和 Yekutieli (2001) 对 1995 的 BH 程序进行了更加保守的调整。对贝叶斯错误发现率而言, $\widehat{\text{FDR}}(z)$ 只涉及分布函数的经验估计, 对统计量是否存在相关不做假设, 仅仅要求经验分布是分布函数的一个优良估计值。进一步的研究和经验表明, 虽然相关会导致控制或估计方法的精度下降, 但是错误发现率对相关具有很好的稳健性。而 Efron (2010) 认为, 大规模数据分析中关于错误发现率所有的问题都不如想象的那样严重, 但除了无效分布 (Null Distribution)。

10.3.7 无效分布

这里所谓的无效分布是指在 H_0 的条件和其他假设下统计量的抽样分布, 例如, 在 $H_0 = 0$ 成立以及正态和等方差等条件下, 两独立样本的 t 检验中统计量 t 具有理论无效分布 $F(v)$, v 为自由度。为了方便, 进行如下的转换:

$$z = \Phi^{-1}(F(t | v)), p = 2\Phi(-|z|), i = 1, \dots, m \quad (10-18)$$

这种转换不只局限于 t 值, 其他统计量如相关系数也可找到类似的转换关系。显然如果 $H_0 = 0$, 则 $z \sim N(0, 1)$, 将 $N(0, 1)$ 称为理论无效分布 (Theoretical Null Distribution, TND)。前文的程序

和估计方法都是依赖于理论无效分布进行的,但是,大规模数据分析中不假思索地利用理论无效分布是否正确呢?

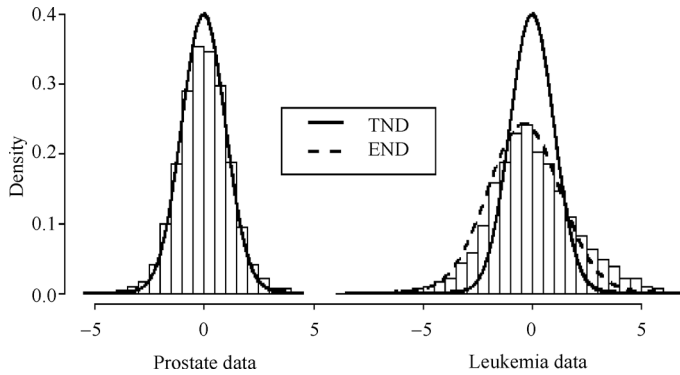


图 10-9 前列腺癌数据和白血病数据 z 值的直方图及无效分布和边际密度曲线

为了说明这一点,图 10-9 给出了两个微阵列数据 z 值的直方图,前者来源于一个关于前列腺癌的数据(Singh et al., 2002),描述了 50 个正常对照和 52 个前列腺癌患者 6033 个基因的表达水平(见后文实例);后者来源于 Golub et al. (1999) 的白血病数据,描述了 72 个白血病患者 7128 个基因的表达水平,其中急性淋巴细胞白血病患者 45 个和急性髓性白血病患者 27 个。医学常识可知,在数以千计的基因中,只有少部分基因的表达水平会存在差异,而绝大部分基因在两组人群中表达一样。基于这种事实,那么处于 0 附近的 z 值应该被合理地认为来自理论无效分布 $N(0,1)$,换句话说,那些绝对值小的 z 值最有可能服从 $N(0,1)$,这种假设称为零假设条件(Zero Assumption, ZA)。图 10-9 中前列腺癌数据 0 附近的 z 值和 $N(0,1)$ 比较吻合,表明理论无效分布是合理的,但白血病数据 0 附近的 z 值明显异于 $N(0,1)$,理论无效分布值得怀疑。然而幸运的是,不像单个假设检验那样只能被动地接受理论无效分布,大规模数据的平行结构再一次使得经验估计无效分布变得可能,称从数据估计得到的无效分布为经验无效分布(Empirical Null Distribution, END)。不同方法被用来估计经验无效分布,这里采用 Efron(2004)提出的中心匹配估计(Central Matching Estimation),思路如下。

(1) 仍然假设经验无效分布为正态分布,但具有不一样的参数: $f_{0ZA} = N(z | \mu, \sigma^2)$ 。零假设条件暗示,如果 f_{0ZA} 为正态,那么处于 0 附近的 z 值应该和 $f(z) = \pi_0 f_{0ZA}$ 一致, $\pi_0 f_{0ZA}$ 具有 z 的二次形式;如何定义“0 附近”仍值得商榷。

(2) 采用 Poisson 回归方法估计边际密度 f , 得到概率密度函数的参数形式 \hat{f} ;

(3) 对 0 附近的 z 值建立 $\log \hat{f}$ 关于 z 的二次多项式方程:

$$\begin{aligned} \log \hat{f}(z) &= \log(\pi_0 f_{0ZA}) = b_0 + b_1 z + b_2 z^2 \\ &= \log \pi_0 - \left[\frac{\mu^2}{\sigma^2} + \log(2\pi\sigma^2) \right] / 2 + \frac{\mu}{\sigma^2} z - \frac{1}{2\sigma^2} z^2 \end{aligned} \quad (10-19)$$

采用最小二乘估计获得系数的估计值 $\hat{b}_0, \hat{b}_1, \hat{b}_2$;

(4) 匹配方程中的系数得到:

$$\begin{aligned} \hat{\mu} &= -\hat{b}_1 / (2\hat{b}_2), \hat{\sigma} = \sqrt{-1 / (2\hat{b}_2)} \\ \hat{\pi}_0 &= \min\{\exp[\hat{b}_0 + 1^2 / 2 + \log(-\pi / \hat{b}_2) / 2], 1\} \end{aligned} \quad (10-20)$$

将区间 $[-0.5, 0.5]$ 作为“0 附近”的定义,即选择在 $[-0.5, 0.5]$ 内的 z 值进行第三步的最小二乘估计。中心匹配估计获得经验无效分布参数为: $\hat{\mu} = -0.379$ 、 $\hat{\sigma} = 1.651$ 和 $\hat{\pi}_0 = 0.929$ 。

图 10-9 中给出了白血病数据的经验无效分布曲线, 对白血病数据如果选择 $N(0,1)$ 作为无效分布, $\pi_0 = 1$, $q = 0.05$, 根据 BH 程序得到 1188 个基因, 选择经验无效分布 $N(-0.379, 1.651)$, 得到 14 个基因, 可见不切实际的理论无效分布会显著地影响大规模数据的分析结果。

10.3.8 软件执行

Benjamini 和 Hochberg 1995 年的论文十分流行的另一个重要原因在于 BH 程序的执行十分简单。这里主要介绍 R 软件中与错误发现率相关的软件包和函数, 表 10-3 给出了部分程序一个概括性的描述, FDR 指基于尾部面积的错误发现率, fdr 指基于单个统计量的局部错误发现率, 参数指该软件要求的输入数据类型, 估计方法指用于估计 FDR 或密度函数的方法, z 代表 z 统计量或其他形式的统计量如 t 值和相关系数 r 等, p 代表根据统计量 z 获得的无效分布尾部面积。

表 10-3 R 软件中介绍的部分与错误发现率有关的软件包

软 件 包	FDR / fdr	参 数	估 计 方 法	主 要 作 者
p.adjust	FDR	p	BH 或 BY	{ stats }
mixfdr	FDR / fdr	z	正态混合	Muralidharan
qvalue	FDR	p	重抽样和样条函数	Storey
multtest	FDR	p	BH	Dudoit
SAGx	FDR / fdr	p	Grenander	Broberg
fdrtool	FDR / fdr	z/p	调整的 Grenander	Strimmer
locfdr	fdr	z	Poisson 回归	Efron

这些软件包的主要区别包括: ①输入参数的不同, 大多数依赖 p 值, 而扩展到 z 值的软件包还具有估计经验无效分布的优势。只有 fdrtool 和 locfdr 对无效分布进行经验估计, 其他软件包都忽略这一问题; ②更重要的区别在它们在估计密度函数和零假设比例 π_0 的方法不同。

对于现代大规模数据分析而言, 错误发现率的出现具有里程碑的意义, Benjamini 和 Hochberg 1995 年的论文也成了过去十几年最具有影响的统计论文之一。错误发现率能够在不同统计哲学观点间建立桥梁, 具有频率统计、贝叶斯和经验贝叶斯及决策论的解释, 它的这种潜在理论价值让统计学家们看到了在大规模数据背景下频率统计、贝叶斯统计和经验贝叶斯在 21 世纪相互融合的希望。

错误发现率的理论还使得统计学家们开始重新审视传统统计中的某些观点, 比如假设检验和估计方法之间的关系, Benjamini (2010) 用一个合成的术语“testimation”来描述这种概念; 其他包括错误覆盖率(False Coverage Rate, FCR), 一个和错误发现率类似的、针对可信区间的概念。错误发现率的理论还使统计学家们注意到, 在大规模数据中传统的理论无效分布可能是不切实际的, 需要更加准确的无效分布以满足实际数据。

总之, 以大规模数据为契机和错误发现率为代表的理论, 为 21 世纪的统计展示了一幅充满希望和挑战的蓝图。

10.4 实例分析和结果报告

本节主要介绍基于错误发现率的微阵列差别表分析, 错误发现率基本已成为大规模多重假设检验的一个基本准则。为了简洁, 本文的实例并不涉及前文的贝叶斯错误发现率和无效分布等方面的内容, 但后面给感兴趣的读者提供了进一步了解的参考文献。下面是一个典型的没有考虑多重假设检验的实际应用例子。

目的用 DNA 微阵列筛选异氟醚麻醉的表达基因。方法 12 只雄性 Wister 大鼠, 分为 2% 异氟醚组(A 组)和对照组(B 组)。A 组用 2% 异氟醚麻醉 1h, B 组直接处死后取脑组织。提取 RNA, 荧光标记, 用有 4096 个位点的 BiostarR-40s 基因芯片, 检测差异表达的基因。结果发现了 26 个差异表达的基因, 其中 17 个基因下调, 9 个基因上调。结论基因芯片检测出 2% 异氟醚麻醉 1h 有 26 个差异表达基因。

统计分析筛选出 $Cy5/Cy3 > 2$ 或 0.5 的数据作为差异表达基因。 $Cy5/Cy3 > 2$ 为差异表达上调, $Cy5/Cy3 < 0.5$ 为差异表达下调。

显然, 文中存在的主要问题是: ①没有考虑多重假设检验所产生的错误率控制问题; ②以比值 $Cy5/Cy3$ 作为差异表达的依据具有一定的主观性。正确的解释和思路见前文和后文。由于不可能得到原始数据, 下面通过另外的实际数据介绍本方法。

10.4.1 前列腺癌微阵列数据实例

微阵列数据一般表示为 $m \times n$ 的矩阵形式, m 表示基因数, n 表示生物样本, 通常只有几个或者几十个。表 10-4 给出了一个关于前列腺癌的微阵列试验数据(Singh et al., 2002), 包括 50 个正常对照和 52 个前列腺癌患者 6033 个基因的表达水平, 以 6033×102 的矩阵排列, 即 $m = 6033, n = 102$ 。

表 10-4 前列腺癌微阵列数据形式

基因编号	正常对照基因值			前列腺癌患者基因值			<i>t</i> 值	<i>P</i> 值
	1	...	50	1	...	52		
gene1	-0.93	...	-1.09	-1.10	...	3.43	1.484	0.1410
gene2	-0.84	...	-0.83	-0.83	...	-1.17	3.670	0.0004
...
gene6032	-0.90	...	1.68	-0.06	...	-0.71	0.104	0.9175
gene6033	-0.25	...	0.44	-0.81	...	-0.16	-0.913	0.3632

10.4.2 *t* 检验和总体错误率

研究者的主要目的是, 哪些基因在正常人和前列腺癌患者中的表达水平不同? 需要识别出这些基因以进一步分析。关键的问题就在于如何在成千上万的候选基因中找出特征基因。两个独立样本的 *t* 检验可用来检验两组基因的平均表达水平是否存在差异。在零假设条件下, 统计量 $t_i (i = 1, 2, \dots, 6033)$ 服从自由度为 $n - 2 = 100$ 的 *t* 分布, 对应的 P_i 为 $\text{Prob}(|T| \geq |t_i|)$ 。表 10-4 中的最后两列给出了每个基因的 *t* 值和 *P* 值, 图 10-10 给出了 *t* 值的直方图和无效分布(Null Distribution, 即自由度为 100 的 *t* 分布)曲线, 0 附近的 *t* 值和无效分布吻合, 但处于两侧的 *t* 值比无效分布更加分散, 暗示这些位置的基因表达可能存在差别。按照 $\alpha = 0.05$ 的检验水准, 有 478 个 $P_i \leq 0.05$, 显然这个结果值得怀疑, 因为即使所有 6033 个基因表达没有差异, 按照 0.05 的水准大约也会获得“有差异”的基因为 $302 (6033 \times 0.05)$ 个, 因此需要对多重假设检验所带来的错误采取控制措施。常用的总体错误率控制程序包括 Bonferroni、Holm、Hochberg 和 Hommel 等方法, 但对前列腺癌数据都只能得到 3 个差别表达基因。总体错误率控制程序没有发现更多基因的原因在于, 对大规模的微阵列数据而言, 总体错误率过于严格, 例如, Bonferroni 的检验水准为 $8.3 \times 10^{-6} = 0.05/6033$, 只有极小的 P_i 所对应的基因表达才能被认为有差别。

10.4.3 错误发现率控制

选择 $\alpha = 0.05$, 按照 BH 程序得到 21 个差别表达基因。图 10-11 给出了 BH 控制的示意图, 直线

过原点, 斜率为 $0.05/6033$, 参考线左侧的 P 值被认为有差别, 此时对应的 P 值为 1.3×10^{-4} , 也即此时的检验水准为 1.3×10^{-4} 。 $\alpha = 0.05$ 的含义在于, 平均而言, 在所有 21 个差别表达基因中大约有 $1(21 \times 0.05)$ 个基因属于错误识别。如果选择 $\alpha = 0.10$ 或 0.20 , 分别得到 60 或 106 个表达差别基因。同理, $\alpha = 0.10$ (或 0.20) 的含义为在 60 (或 106) 个差别基因中约有 6 (或 21) 个属于错误发现。

10.4.4 错误发现率估计

BH 程序将错误发现率控制在一个预先选择的范围内, 一个自然的问题是, 例如, 以 $|t_i| \geq 3$ 作为拒绝域得到 105 个基因, 那么对应的错误发现率是多少呢? 这一过程和上述的 BH 控制相反, 称为错误发现率估计。如果所有 6033 个基因表达是无差异的, 那么理论上分布在区间 $(-\infty, -3] \cup [3, \infty)$ 的 t 值个数约为 $21 = 6033 \times 2 \times F(-3, 100)$, $F(x, 100)$ 表示自由度为 100 的 t 分布累计概率函数, $F(-3, 100) = 1.7 \times 10^{-3}$, 对应的错误发现率约为 $0.20 = 21/105$ 。

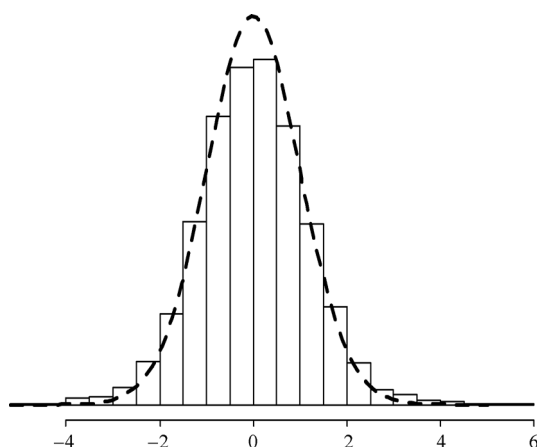


图 10-10 前列腺癌微阵列数据 t 值的直方图和无效分布

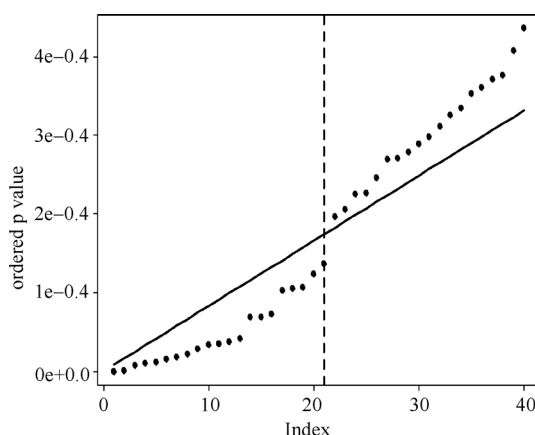


图 10-11 前列腺癌微阵列数据 BH 控制的示意图

10.4.5 结果报告

上述的过程给出了具体的分析过程和思路, 读者可根据实际情况选择报告相应内容。最后, 给出一个简洁的结果陈述。

方法: 6033 个基因执行两独立样本的 t 检验, 采用错误发现率控制多重假设检验中的错误率。
结果: 将错误发现率控制在 0.05 以下时, 通过 BH 程序得到 21 个差别表达基因; 如以 $|t_i| \geq 3$ 作为拒绝域得到 105 个基因, 对应的错误发现率估计值为 0.20。

Methods: 6033 genes are compared through independent-sample t -test, using false discovery rate to control the error rate because of multiple hypothesis testing. **Results:** 21 differential expression genes are identified by the BH procedure while false discovery rate is controlled below 0.05; if rejection regions are $|t_i| \geq 3$, then 105 genes are regarded as differential expression, the corresponding false discovery rate is estimated at a value of 0.20.

10.4.6 前列腺癌 FDR 分析的程序

(1) R 程序

```
x = read.table("d:\\data\\prostatedatasam.txt", sep = "\t")
attach(x)
```

#读取数据

```

x=as.matrix(x)
m=nrow(x)
n1=50
n2=52
n=ncol(x)
y=c(rep(1, 50), rep(2, 52))
tt=rep(NA, m)                                     #建立一个 t 检验的程序
ttest = function(x, y)
{
  mean1 = apply(x[, y==1], 1, mean)
  mean2 = apply(x[, y==2], 1, mean)
  v1 = apply(x[, y==1], 1, var)
  v2 = apply(x[, y==2], 1, var)
  s1 = sqrt(v1/n1)
  s2 = sqrt(v2/n2)
  tt = (mean2-mean1)/(sqrt(s1^2 + s2^2))
  return(tt=tt)
}
tt = ttest(x, y)                                   #计算 P 值
p=pt(-abs(tt),100)* 2
tvalue=tt
par(mfrow=c(2,2),mgp=c(1.2,0.3,0),tcl=-0.2,font.axis=2,font.lab=2)
hist(tvalue, breaks=20, probability=T, ylim=c(0,0.4)) #图 10-7
curve(dt(x, 100), xlim=c(-5.7, 5), add=T, lty=2, lwd=3, col="red")
p=sort(p)                                           #图 10-8
plot(p[1:40], pch=20, col="red", ylab="ordered p value")
lines(seq(1:40), seq(1:40)* 0.05/6033, lwd=2, col="blue")
abline(v=21, lty=2, lwd=2)
lines(density(tvalue), lwd=2, col="blue")          #FWER 控制程序
sum(p.adjust(p,"bonferroni") <=0.05)
sum(p.adjust(p,"holm") <=0.05)
sum(p.adjust(p,"hochberg") <=0.05)
sum(p.adjust(p,"hommel") <=0.05)
sum(p.adjust(p,"BH") <=0.05)
sum(p.adjust(p,"BH") <=0.1)
sum(p.adjust(p,"BH") <=0.2)
sum(abs(tvalue) >=3)

```

需要注意的是：①R 程序中的英文是区分大小写的。②程序中可直接调用 R 的内置 t 检验程序 `t.test`，但此处自定义了一个两独立样本的 t 检验程序，即文中 `ttest` 函数。

(2) SAS 程序

```

options ls=80 ps=60;                               #定义 SAS 环境读取数据
data prostate;
set "d:\data\prostate\data.sas7bdat";              #转换数据的格式
proc transpose data=prostate
out=prostate\data prefix=gene;
var sub1-sub102;

#定义分组变量

data prostate;
set prostate;
group=(_n_>50);
#调用 SAS 的 multtest 过程,选择 fdr 参数

```



```

proc multtest data=prostate seed=37081fdr out=prostatevalue ;
class group;
test mean (gene1-gene6033);
Contrast "means" -1 1;
run;

proc sort data=prostatevalue;          #获得经过 fdr 校正后的 P 值
by raw_p;

data prostatevalue;
set prostatevalue;
T=_value_/_se_;
proc print data=prostatevalue;
run;

```

参 考 文 献

- [1] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 1995, 57(1): 289-300.
- [2] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 2001, 29(4): 1165-1188.
- [3] Benjamini Y. Comment: microarrays, empirical Bayes and the two-groups model. *Statistical Science*, 2008, 23(1): 23-28.
- [4] Benjamini Y. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B*. 2010, 72: 405-416.
- [5] Bretz F, Hothorn T, Westfall P. *Multiple Comparisons Using R*. London: Chapman & Hall, 2010.
- [6] Dudoit S, Shaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 2003, 18(1): 71-103.
- [7] Dudoit S, van der Laan MJ. *Multiple Testing Procedures and Applications to Genomics*. New York, Springer, 2007.
- [8] Efron B. Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, 1979, 7(1): 1-26.
- [9] Efron B, Tibshirani R. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993.
- [10] Efron B, Tibshirani R, Storey J D, et al. Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*, 2001, 96(456): 1151-1160.
- [11] Efron B. Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, 2005, 100(469): 1-5.
- [12] Efron B. Doing thousands of hypothesis tests at the same time. *Metron International Journal of Statistics*, 2007.
- [13] Efron B. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. New York: Cambridge University Press, 2010.
- [14] Efron B. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, 2004, 99: 96-104.
- [15] Efron B. Local false discovery rates. Department of Statistics Technical Report 234, Stanford University, 2005. Available at <http://stat.stanford.edu/~brad/papers/>.
- [16] Efron B. Microarrays, empirical Bayes, and the two-groups model. *Statistical Science*, 2008, 23(1): 1-22.
- [17] Efron B. Robbins, empirical Bayes, and microarrays. *Annals of Statistics*, 2003, 31(2): 366-378.
- [18] Fan JQ, Li R. Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery. *Pro-*

- ceedings of the International Congress of Mathematicians (Sanz-Sole M, Soria J, Varona JL, Verdera J, eds.), Vol. III, European Mathematical Society, Zurich, 595-622.
- [19] Golub TR, Slonim DK, Tamayo P, Huard C, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 1999, 286: 531-537.
- [20] Good PI. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. New York: Springer, 2005.
- [21] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. New York: Springer-Verlag, 2009.
- [22] Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *New England Journal*, 2001, 344(8): 539-548.
- [23] Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 1988, 75: 800-803.
- [24] Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 1979, 6: 65-70.
- [25] Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 1988, 75: 383 - 386.
- [26] Johnstone IM, Titterton DM. Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A: Physical, Mathematical and Engineering Sciences*, 2009, 367: 4237-4253.
- [27] Lazar N. *The Statistical Analysis of Functional MRI Data*. New York: Springer, 2008.
- [28] Lehmann EL, Romano J. *Testing Statistical Hypotheses*, 3rd edition. New York: Springer, 2005.
- [29] McConnell P, Lin SM, Hurban P, Eds. *Methods of Microarray Data Analysis V*. New York: Springer, 2007.
- [30] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. URL <http://www.R-project.org>.
- [31] Schena M. *Microarray Analysis*. New York: John Wiley & Sons, 2003.
- [32] Shaffer JP. Multiple Hypothesis Testing. *Annual Review of Psychology*, 1995, 46: 561-584.
- [33] Shao J. *Mathematical Statistics*, 2nd Edition. New York: Springer, 2005.
- [34] Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 2002, 1(2): 203-209.
- [35] Storey JD, Tibshirani R. Statistical significance for genome-wide experiments. *Proceeding of the National Academy of Sciences*, 2003, 100(16): 9440-9445.
- [36] Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, 2003, 31: 2013-2035.
- [37] Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B*, 2004, 66: 187-205.
- [38] Storey JD, Tibshirani R. Statistical significance for genome-wide experiments. *Proceeding of the National Academy of Sciences*, 2003, 100: 9440-9445.
- [39] Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 2002, 64: 479-498.
- [40] Storey JD. False discovery rates. In *International Encyclopedia of Statistical Science*, 2010.
- [41] Strimmer K. A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 2008, 3: 1.
- [42] Turnbull B. Optimal Estimation of False Discovery Rates. Available at <http://www.stanford.edu/~bkatz/>.
- [43] Tusher V, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 2001, 98: 5116-5121.
- [44] Westfall PH, Young SS. *Resampling-Based Multiple Testing Examples and Methods for P-Value Adjustment*. New York: John Wiley & Sons, 1993.

- [45] 陈希孺. 数理统计学简史. 长沙: 湖南教育出版社, 2003.
- [46] 顾健腾, 陶国才. 蛋白质芯片与全身麻醉机制. 国外医学: 麻醉学与复苏分册, 2005, 4: 203-205.
- [47] 郭明明, 余之刚, 马忠兵等. 基因芯片技术筛选乳腺癌相关差异表达基因. 山东大学学报(医学版), 2011, 49(7): 910-990.
- [48] 胡良平, 李子健. 医学统计学基础与典型错误辨析. 北京: 军事医学科学出版社, 2003.
- [49] 李明强, 曾邦雄. 基因表达谱芯片用于检测脓毒症小鼠肺组织基因表达的改变. 中华麻醉学杂志, 2003, 9: 667-670.
- [50] 陆守曾, 陈峰. 医学统计学(第2版). 北京: 中国统计出版社, 2007.
- [51] 茆诗松, 王静龙. 高等数理统计. 北京: 高等教育出版社, 2006.
- [52] 苏殿三, 王祥瑞. 急性肺损伤易感基因的研究进展. 国外医学: 麻醉学与复苏分册, 2005, 6: 330-333.
- [53] 苏良军. 高等数理统计. 北京: 北京大学出版社, 2007.
- [54] 孙振球. 医学统计学(第2版). 北京: 人民卫生出版社, 2005.
- [55] 王振猛, 李泉. 基因芯片与麻醉学. 国外医学: 麻醉学与复苏分册, 2002, 6: 321-324.
- [56] 吴财能, 江鹏, 曹莹等. 舒芬太尼预处理对糖尿病大鼠烫伤后心肌损伤的影响. 临床麻醉学杂志, 2012, 2: 175-177.
- [57] 荀鹏程, 赵杨, 易洪刚等. Permutation Test 在假设检验中的应用. 数理统计与管理, 2006, 25(5): 616-621.
- [58] 周志强, 程璐, 杨建军等. DNA 微阵列检测异氟醚麻醉的基因表达. 临床麻醉学杂志, 2006, 22(5): 369-371.
- [59] 邹定全, 常业恬. 基因芯片技术在心肌缺血再灌注损伤研究中的应用. 国外医学: 麻醉学与复苏分册, 2005, 4: 228-231.

第 11 章 科研设计方法及要点的把握

11.1 科研设计概述

11.1.1 科研设计的概念

何为科研设计？顾名思义，就是进行一项科学研究之前的安排。进行任何一项科学研究，都有一个比较明确的研究目的，科研设计是为了尽可能圆满地完成事先确定的研究目的所做的一切考虑和安排。具体来说，包括研究者对拟研究问题的认识，拟进行何种试验或观察，具体做法如何，预期的目标是什么，用多长时间来完成，需要给予哪些帮助和支持等。

11.1.2 科研设计整体架构

科研设计涵盖面非常广，其基本内容概括起来如图 11-1 所示。

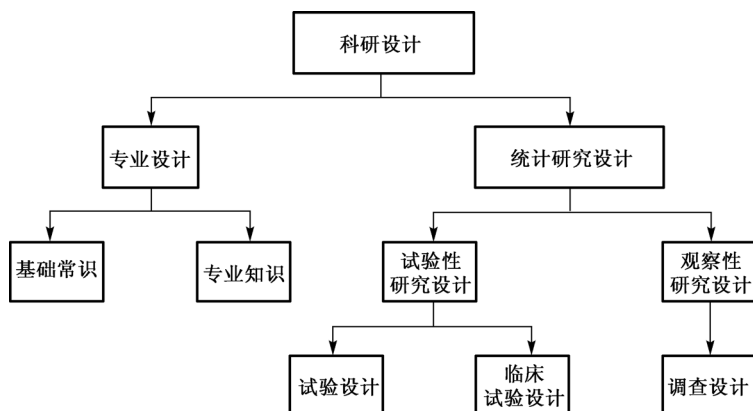


图 11-1 科研设计框图

由图 11-1 可知，科研设计由专业设计和统计研究设计两部分组成。

所谓专业设计，就是从专业角度考虑，要使一项科研工作达到事先确定的目标需要做哪些必要的准备和安排。在这些内容中，必不可少地会包含一些看似微不足道但却又是必不可少的内容，这些内容被称为“基本常识”；除此之外，就是与拟研究问题密不可分却又比较深入的内容，这些内容被称为“专业知识”。

统计研究设计，则是在科学、周密、严谨地考虑问题的基础上，而制订合理的、完善的工作计划或研究方案。统计研究设计的类型通常包括“观察性研究设计和试验性研究设计”两大类，其中观察性研究设计可以细分为横断面研究设计、队列研究设计、病例 - 对照研究设计、混合研究设计；而试验性研究设计可分为单因素 k 水平设计、配伍组设计、交叉设计、拉丁方设计、析因设计、重复测量设计、正交设计、均匀设计等。

11.2 科研设计要点

11.2.1 基本常识与专业知识在科研设计中的作用

1. 为什么专业设计中要包括基本常识

有些东西是必需的,或不假思索就应当那样做的,称为“基本常识”。例如,人们要煮米饭,就必须在米中加水;若是蒸米饭,就必须在锅内放入足量的水。又例如,通常中国人一提到吃饭,就会想到要用碗和筷子;而西方人就会想到要用盘子、刀子和叉子。再例如,研究骨质疏松症对人体造成的伤害有多么严重时,必须排除年龄、性别、锻炼情况、钙的摄入和吸收情况、接受阳光照射情况等重要非试验因素的影响。若受试者都是女性时,不仅要注意年龄,更要注意“是否绝经”、“绝经时间长短”和“此激素水平高低”等对观测结果所造成的不同影响。在进行此类科研课题的试验设计时,有意识地使上述提及的重要非试验因素的影响在骨质疏松组与非骨质疏松组保持均衡一致(即具有可比性),所得出的结论才具有令人相信的基础。这充分说明:灵活而且恰当地运用“基本常识”在专业设计中起着举足轻重的作用。

2. 基本常识对于科研设计为何如此重要

有人认为,在科研设计中,基本常识可有可无,根本不应将其列入专业设计之中。其实,这是很糊涂的观点。盘点一下全国或全世界的生物医学科研课题,其中有相当大比例的课题的科研设计是有问题的,在这些问题中,约有七八成违背了基本常识而导致错误发生。例如,有人研究得了骨质疏松症之后对人的身体造成的损害,选择 30 名 65 岁以上患了严重骨质疏松症的妇女作为病例组,又选了 30 名 25~35 岁未患骨质疏松症的妇女作为对照组,然后,从这两组受试对象身上检测多项定量和定性指标,用以反映她们的身体健康情况。结果发现,病例组妇女各项身体指标都比对照组妇女差。结论是:骨质疏松症给妇女的身体健康带来了很坏的影响。这个结论肯定是很正确的,不做此项研究都可得出这样的结论,现在做了研究,由于其科研设计中犯了常识性错误,反倒使其原本正确的结论变得让人很怀疑。因为病例组妇女不仅患了骨质疏松症,而且年龄很大,已绝经,多方面的不利条件,使她们无法与对照组妇女相比,岂能把最终的糟糕局面完全怪罪于骨质疏松症!

任何忽视基本常识的科研课题,要么得出难以令人置信的原本正确的结论,要么得到经不起时间和实践检验的结论。

3. 基本常识与专业知识的区别是什么

可以这样说,基本常识是专业知识中的 ABC,而专业知识则是较高深、较专门的知识 and 技能。例如,高楼大厦失火了,室内的人不要跳楼逃生,要尽快用湿毛巾盖住口和鼻、顶着湿被子离开着火的房间,这属于基本常识;而如何有效地抢救人员和扑灭大火,则需要掌握消防知识和拥有消防器材的专业人才,这主要靠专业知识和技能。又例如,有人中毒了,尽快将患者送往救治能力强的医院,属于基本常识;而如何快速诊断是什么东西引起的中毒、如何尽快解救中毒患者,则需要有相关专业知识和临床经验的医护人员。

11.2.2 统计研究设计及其内在关系

1. 统计研究设计的共性

虽然统计研究设计由试验设计、临床试验设计和调查设计组成,但其实它们之间有很多共同的

内容。第一,目标相同,都是为了更好地完成事先确定的研究目标而从统计学角度所做的考虑、计划和安排,都希望能透过事物的现象看清其本质。第二,核心内容相同,一般都不可避免地要涉及“三要素”、“四原则”、“设计类型”和“质量控制”。第三,在科研工作具体实施之前,都应对未来的试验或调查过程中可能出现的意想不到的困难、干扰和影响结果正确性的非试验因素,有足够的警觉和必要的防范措施(可简称过程中的质量控制),并在试验或调查全过程中实时监控与纠偏,以确保研究方案的顺利实施和研究结果的正确性与稳定性。

2. 统计研究设计的个性

由于研究的场合、对象和方法等不完全相同,试验设计、临床试验设计和调查设计又具有各自的特性,称其为“个性”。此处将它们各自的个性分述如下。

试验设计:一般指在试验室进行的小规模试验研究所对应的设计。其受试对象通常是动物或样品,各试验因素可由研究者根据基本常识和专业常识选定,非试验因素也比较容易控制,因此,若采用恰当的试验设计类型,可以同时考察很多试验因素及其交互作用对观测结果的影响大小,可以将重要非试验因素的干扰和影响控制在最低水平,过程中的质量可以控制得很严,组间的可比性好,数据的说服力强,所获得的结果准确度高,重现性好,一般来说,所得结论更能经得起时间与实践的检验。

临床试验设计:一般来说,其受试对象是正常人和病人。研究者必然要面对“伦理道德和受试者依从性”两大难题的挑战。怎样设置合理的对照组、如何严格遵守伦理道德、如何选择和剔除受试者、如何提高受试者的依从性、如何控制人为因素(特别是心理因素)引起的偏性等问题的合理解决,是临床试验设计的最重要内容。

调查设计:在没有任何人为干预的前提下,对客观存在的事物、现象或受试对象进行被动的观察,要根据拟完成的调查任务,结合基本常识、专业知识和统计学知识,把现场可能碰到的各种问题尽可能考虑周到,以免在调查结束后,面对“漏项和缺项无法弥补”所带来的尴尬,甚至导致调查研究前功尽弃的悲惨结局。对于一些敏感性调查问题,应采取切实可行的技术方法,以提高调查结果的可信度。调查表内容的设置是至关重要的,既不能过于简单或笼统(影响调查质量),又不能过于复杂或细致(影响调查表的回收率)。

11.2.3 完善的科研设计的标志以及获得完善科研设计方案的途径

1. 完善的科研设计的标志

在科研课题获得批准之后且在课题正式实施之前,应制定出科学完善的科研设计方案。一个完善的科研设计的标志是什么呢?

(1)人力、物力、财力和时间满足设计要求。也就是说,申请书上计划的科研条件,在实际科研环境中均能一一兑现。例如,检测某种物质需要一种昂贵的仪器设备,而本课题组已经拥有或可以借到或可以请他人代为检测,若国内没有一个单位拥有,则属于“空头支票”,不具备完成科研课题的基本条件;又如,完成此科研课题需要10个专业领域的专家参与,而现有研究人员仅涉及6个专业领域,另外4个专业领域的研究人员目前几乎不可能得到,这属于人力上达不到设计要求;再如,要想完成某项难度较大的科研课题,至少需要5年时间,而申请时考虑不周,计划3年完成,这属于时间上不符合设计要求,事先夸下了“海口”,而无力实现。

(2)试验设计的“三要素”和“四原则”均符合专业和统计学要求。这方面的内容很多,将在本章后续部分进行详细讨论。这里仅举一例,例如,拟使用某种动物制造一种用于试验的动物模型,但这种动物很特殊,根据目前的条件尚无法获得,这就属于三要素中的“受试对象”不符合设计要求。

(3)重要的试验因素和观测指标没有遗漏,并做了合理安排。试验因素和观测指标属于三要素中

的两个重要要素，再一次特别提出来，是为了强调其重要性。因为在很多科研课题的设计中，研究者常将某些重要的试验因素和观测指标遗漏掉，导致科研结果回答不了需要回答的问题，甚至很容易得出错误的结论。

(4)重要非试验因素都得到了很有效的控制。很多科研课题经不起时间和实践的检验，问题的症结就在于重要的非试验因素未得到有效的控制，他们“污染”了试验结果，从而得出了不够正确的结论。

(5)研究过程中可能出现的各种情况都已考虑在内，并有相应的对策。这一点也是至关重要的。例如，在新药临床试验研究中，若中途退出的受试者人数较多，可能会导致整个临床试验前功尽弃，一旦发现可能出现较高比例的受试者退出临床试验，应预先想好对策，尽可能降低退出比例；再如，某些受试者的某些测定指标的数值明显偏高或偏低，应事先对此种可能的异常现象有所警惕，必要时，应对所有受试者进行多次重复观测，用每位受试者的平均值取代其测定指标的数值，以降低个别受试者在个别时间点上出现难以处理的异常值。有时，极少数异常值可能会颠覆整个研究结论。

(6)对操作方法、试验数据的收集、整理、分析等均有一套明确的规定和方法。尤其是有多个单位多位研究者合作的研究课题，统一规定试验操作规程、统一培训、仪器和试剂统一质量指标和校准、有规范化且符合试验设计要求的收集和整理资料的表格，对获得准确可靠试验资料是极其重要的；有了数据后，将由谁来进行数据处理、将采用哪些统计分析方法，事先都应做到心中有数，不要等到试验都做完了，再来考虑请谁做数据分析。因为到那时，很可能这个科研课题的数据已无分析的价值了。什么原因呢？试想，若科研设计有严重错误或质量控制有问题或收集资料的方法不正确，都可能得到包含很多错误信息的资料，无论用何等高级的统计分析方法去分析，其结论都是荒谬的。也就是说，统计分析人员应参与开始的科研设计，应在科研过程中参与质量控制，应协助研究人员确保科研资料准确完整。

2. 如何获得完善的科研设计方案

严格地说，十分完善的科研设计方案是很难制定出来的，因为有许多事情是很难事先考虑周到的。但是，如果设计者专业知识和统计学知识都十分丰富，又舍得花时间认真思考、反复琢磨，设计方案是可以制定得相当完美的。然而，任何人的知识和经验都是十分有限的，有必要多集中一些他人的智慧。主要是两方面的人，其一，知识面宽广的同行专家，请同行专家帮助修改完善科研设计方案，是非常有价值的一项措施；其二，经验丰富的统计学专家，统计学专家往往比一般实际工作者对设计方案中存在的瑕疵具有更敏锐的洞察力。所以，一个设计方案，若经设计者深思熟虑，又经同行专家精雕细琢，再经过统计学专家明察秋毫与画龙点睛，有望使其日臻完善。

11.3 试验设计要点

以动物或样品为受试对象的试验研究通常在试验室内完成，在确定试验目标之后，应制订科学完善的试验设计方案，以便试验者在其指导下顺利而又成功地完成试验，获得预期的结果。本节将介绍试验设计方案中的要点，内容包括试验设计三要素、四原则、设计类型和质量控制。

11.3.1 广义、狭义及简易型试验设计分别是什么

1. 广义型试验设计

为成功达到研究目的所做的一切考虑和安排被称为广义型试验设计。例如，为了评价某药物A对某病的有效性与安全性。广义型试验设计至少需要考虑以下5方面的问题。

第一,参加研究的组织者与组织单位、研究者与研究单位及职责分工与经费分配。一个大型的试验研究课题,需要有一个负责单位和个人牵头组织,形成必要的组织机构;招募必要的参加研究单位和个人,共同完成此项试验研究之重任;在试验研究的全过程中,将涉及许多环节,各环节的职责是什么,由谁来承担,应达到什么样的质量标准,这些可称为“职责分工及标准制定”;应根据研究任务的难度、技术含量、劳动强度及将付出的时间与人力,分配相应的科研经费,以确保课题研究顺利进行下去。

第二,试验研究目标和试验方案的制订、完善与落实。应由项目负责人组建一个课题研究领导小组,负责制订试验研究目标、制订试验设计方案的初稿,并召集有关多种学科的专家对试验研究方案进行论证,认真修改,使其逐渐完善;还应制订出如何确保试验研究方案得以严格落实的措施。

第三,完成试验研究所需硬件与软件条件的提出与准备。硬件通常指完成特定试验研究所需要的场地、仪器设备、药品、试剂等;软件通常指掌握相应专业知识的技术力量(专业人才和技术顾问)、技术资料与方法、统计或计算机软件等。应由课题领导小组提出完成此项试验研究所需要的硬件与软件条件,并确保在试验研究之前落实到位。这是课题研究成功的必备基础。

第四,试验过程中的质量控制、监督与管理。在试验设计阶段,就应提前想到试验过程中可能出现的“产品”质量问题,对每种可能导致偏离预期结果的影响因素,都应有相应的处理预案,对试验过程中出现的任何质量问题都能进行适时控制、监督和管理,确保试验研究过程中的质量始终处于严格掌控之中。

第五,试验结束后的验收、评估和成果上报。这些内容似乎不应该在试验设计方案中体现,但从“科学完善”的要求来看,应将其纳入之中,以显示做事“善始善终”,而不是“虎头蛇尾”。从高标准、严要求的角度来看,在试验设计阶段,就对此项试验研究将来可能获得的成果档次有一个比较符合实际的预期,也不失为对项目参加者鼓舞士气、拼搏进取的动员令。

2. 狭义型试验设计

在试验设计方案中,仅围绕如何做好与“三要素”、“四原则”和“设计类型”有关的各项内容的考虑与安排。这些内容将在下一小节中介绍,此处从略。

3. 简易型试验设计

在试验设计方案中,仅考虑与试验因素和重要非试验因素的安排有关的问题。具体来说,就是如何安排试验因素和重要的非试验因素,常被称为“试验设计类型”的确定和设计架构的展示。这些内容将在下一小节中介绍,此处从略。

11.3.2 试验设计三要素及把握的要领

1. 何为试验设计三要素

在进行试验研究之前,需要根据研究目的制订一个完善的计划,即试验设计方案,以便指导试验顺利进行下去,并确保以较少的人力、物力、财力和时间得到准确可靠的试验结果。试验设计方案的核心内容为试验设计的三要素、四原则和设计类型。任何一项试验研究都包括受试对象、影响因素和观测指标三个部分,故称它们为试验设计的三要素。

2. 把握受试对象的要领

1) 受试对象的选取

(1)选取受试对象的一般原则。受试对象选择得当,能够为试验成功创造有利条件。根据研究目的的不同,受试对象可以是人、动物或植物,也可以是某个器官、组织、细胞、亚细胞或血

清等生物材料。受试对象的选择应以试验研究的目的为依据,并结合专业知识来确定。医学研究一般需要先进行动物试验,在确定无严重毒副作用且具有较好疗效和安全性的条件下再进行临床试验。

根据试验研究的目的和具体情况,应结合专业知识选择合适的受试对象。从统计学角度看,更多场合下是应当结合具体问题制定出“纳入和剔除”受试对象的标准,以减少或消除重要非试验因素对试验结果的干扰和影响。受试对象的同质性必须予以高度重视。在药物的临床试验中,还必须考虑到“伦理道德”、“尊重人权”和“提高受试者依存性”等问题。

无论受试对象是动物还是人,首先应满足以下三个条件:①对试验因素敏感;如研究某药物对高血压的治疗效果,常选用Ⅰ、Ⅱ期高血压患者作为受试对象,因为Ⅲ期高血压患者对药物不够敏感;②反应必须稳定;③为使研究结果具有普遍性和推广价值,须保证受试对象的同质性和代表性。在进行临床试验时,选择的病例应体现这种疾病的特点,如果研究的某种疾病好发于老人,而选择的研究对象却是青年人,试验结果就难以说明问题。此外,不同的试验,还有一些特殊的纳入和排除标准。

(2) 选择受试对象的基本要求。选择受试对象时应着重考虑下列基本要求。

①敏感性:即受试对象对被施加的处理因素应有较高的敏感性,容易显示效应。②特异性:即受试对象对被施加的处理因素应有较强的特异性,排除非处理因素干扰。③稳定性:即受试对象对被施加的处理因素的反应有较大的稳定性,减少误差。④同质性:研究对象的种属、生物学特性及其他条件保持均衡。⑤依从性:受试对象接受处理因素的合作程度。⑥经济性:指受试对象容易找到,费用低。

2) 受试对象的种类

根据受试对象的特征和属性可将其粗分为生物体和非生物体两大类,生物体又可粗分为人体和非人体两类,其中人体又可细分为患者和非患者,非人体又常细分为动物、植物,另外,还可以是人体的离体标本或细胞等。不同的试验目的、试验因素、试验的技术水平需要选取不同的受试对象。

(1) 受试对象:动物。根据课题研究的目的、内容、水平,选择相匹配的试验动物。需遵循以下原则:①选用解剖、生理特点符合试验目的要求的试验动物。②尽量选择研究对象的功能、代谢、结构及疾病性质与人类相似的动物。③根据试验动物不同品种、品系的特点选择动物。④必要的预试验有助于选择与课题相适应的试验动物。⑤试验动物的选择和应用要注意有关国际规范和动物福利。⑥按 3R 要求选择试验动物,3R 是指 Reduction(减少)、Replacement(替代)和 Refinement(优化)。“减少”是指减少试验用的动物和试验的次数;“替代”是指尽可能采用可以替代试验动物的替代物,如用细胞、血清等离体标本;“优化”是指对待试验动物和动物试验工作应做到尽善尽美。⑦尽量选用容易获得、价格便宜和饲养经济的动物。⑧年龄和体重:一般均选用成年动物来进行试验,但不同试验对年龄要求不尽相同,需根据课题的内容而定;试验动物年龄与体重一般呈正相关,一般来说,选择的试验动物年龄、体重应尽可能一致,相差不得超过 10%。⑨性别:如无特殊需要,一般宜选用雌雄各半。⑩生理状态:试验动物应证明确实健康,如无特殊要求,处于怀孕、授乳期的雌性动物一般不宜采用。以某种动物为受试对象的实例见【例 11-1】。

【例 11-1】《悬吊模拟失重及解悬吊对大鼠骨密度及生物力学的影响》,20 只 wistar 雄性大鼠随机分成空白组和模型组。模型组尾部悬吊 14d,解悬吊后继续饲养 14d,空白组则正常饲养 28d。试验第 28 天处死大鼠,检测右侧股骨及 L₄椎体 BMD 及生物力学强度,差异有统计学意义,资料见表 11-1。不难看出,在表 11-1 中,受试对象为“20 只 wistar 雄性大鼠”。

表 11-1 两组动物试验第 28 天离体股骨、腰椎骨密度的变化 ($\bar{x} \pm s$)

组 别	鼠数	骨密度 (g/cm^2)	
		股骨	L ₄ 腰椎
空白组	10	0.1535 \pm 0.0102	0.1438 \pm 0.0079
模型组	10	0.1353 \pm 0.0112	0.1289 \pm 0.0069

受试者诊断应该有统一的诊断标准，最好是国际公认的标准，或被国内同行一致认可的标准，若没有统一的标准，可以自己设定，必须尽可能地采用客观的标准，在操作时便于明确诊断。③依从性好，首先由于心理、社会、经济等多方面原因，患者可能出现忘记服药、中途退出试验或出现换组，应尽可能减少依从性差的受试者人数；其次由于病情急剧恶化或存在难以忍受的副作用，必须中途退出试验，这些不依从性表现必然干扰试验计划的实施。因此，必须充分关心体贴患者，做好思想工作，使患者建立充分信任与依赖医务人员心理状态，从而提高依从性，并且应当控制试验时间，试验时间过长往往依从性降低。④被选择的对象应该从试验研究中可能受益。⑤已知试验对其有害的人群，不应作为试验对象，例如有消化道出血史者不应作为抗炎药物的试验对象。⑥对于一些受试对象患有可能会影响试验结果的疾病，或这些疾病本身并不影响试验结果，但治疗这些疾病所用的药物或措施可能影响试验结果，这些病例必须予以排除。以某病患者为受试对象的实例见【例 11-2】。

【例 11-2】《旋转扳法和穴位推拿治疗青年颈性眩晕的随机对照试验》，76 例表现为椎基底动脉异常高流速状态的青年颈性眩晕患者随机分为旋转扳法组 (A 组，38 例) 和穴位推拿组 (B 组，38 例)，分别施加相应的治疗手法。经颅超声多普勒 (TCD) 检测治疗前后椎基底动脉流速并观察流速改变和临床疗效，资料见表 11-2。

(3) 受试对象：离体标本或细胞。

在某些试验研究中，若可以使用离体标本、血清或细胞作为受试对象，应根据研究目的，事先也应制定相应的纳入与剔除标准，这是确保试验数据具有可比性的重要前提。用离体标本的实例见【例 11-3】。

【例 11-3】《前下后上型齿突骨折螺钉内固定的生物力学研究》，收集 20 具新鲜成人上颈椎标本，分为 3 组，模拟齿突 II 型骨折模型，骨折线为前上后下，截骨角度分别为 0°6 组，17°8 组，25°6 组，并进行齿突螺钉固定。进行旋入力矩、最大轴向拔出力及刚度测试。各组最大旋入力矩、抗拔出力差异无统计学意义，资料见表 11-3。

表 11-2 两组治疗后 6 个月疗效比较 (例)

组 别	例 数				
	治愈	显效	有效	无效	合计
旋转扳法组	13	7	8	10	38
穴位推拿组	20	12	3	3	38

(2) 受试对象：某病患者。

受试对象通常为患者或正常人群，应注意其性别、年龄、病情、病程、民族、职业、文化程度、经济状况和心理素质。需遵循以下原则：①受试者同意参加试验，签署知情同意书，并符合伦理道德原则。②诊断明确，对

表 11-3 不同截骨角度与各测量参数值 ($\bar{x} \pm s$)

组 别	最大旋入力矩 ($\text{N} \cdot \text{m}$)	骨折移位 (mm)
0°截骨组	4.53 \pm 2.59	0.70 \pm 0.61
17°截骨组	6.18 \pm 2.22	1.61 \pm 0.64
25°截骨组	4.47 \pm 1.77	2.88 \pm 0.80

3) 受试对象的数量

受试对象的数量通常指试验研究中总共需要多少样本含量，也称为样本大小，在统计学上称为“样本大小估计问题”。值得注意的是，若试验涉及的因素很多，根据因素的水平组合会形成很多小组，每个小组中的受试对象的数量不能太小！样本大小估计是一个比较复杂的问题，它涉及试验设计的类型、比较的类型 (如差异性检验、非劣效检验、等效性检验或优效性检验)、观测指标的性质 (定量与定性)、有关先验知识 (如均值和标准差大约为多少) 和对结果精确度 (允许估计或分析

的结果中犯第一类和第二类错误的概率大小)的要求。比较好的做法是根据具体情况利用相应的计算公式估算出样本含量。

4) 实际工作者在选取受试对象时常犯的错误

(1) 在以动物为试验研究的场合中,常盲目选取某种受试对象,其结果易于导致试验失败。在制作不同疾病的动物模型时,必须依据专业知识和以往的经验,选择最合适的动物作为受试对象。否则,试验很容易失败。例如,制作糖尿病动物模型适合选用自发性高血压大鼠;制作高血脂或动脉粥样硬化动物模型适合选用日本大耳白兔。又如,做染色体畸变试验与分析时,由射线照射引起染色体畸变,观察指标为“双着色粒、染色体环”的发生率,发生率越高,表明剂量越大,在一定剂量范围内,剂量与效应之间呈线性关系,关键在于选用什么细胞作为受试对象,是选外周血淋巴细胞还是选骨髓细胞?国际上公认用“外周血淋巴细胞”,因为淋巴细胞相对比较一致、均匀,且好采样。后来,有人为了“创新”,提出用“骨髓细胞”来培养。众所周知,骨髓细胞中有红细胞系统、粒细胞系统、血小板系统,还有淋巴细胞和其他细胞,不仅细胞系统多、杂,而且还同时存在处于幼、早、中、晚期的细胞,因此,这样的细胞体系不能很好地反映剂量与效应之间的线性关系。事实证明,不宜用这样的细胞体系来做此类试验研究的受试对象!

(2) 在以人为受试对象的试验研究中,常忽视伦理道德问题,选择不恰当的对照组,或在知情同意书上使用受试对象不易理解的表述方法或语言。

3. 把握影响因素的要领

1) 何为影响因素

何为因素与水平?因素通常是指性质不同的不同试验条件的总称,而性质不同的不同试验条件又被称为该试验因素的不同水平。如需在一定温度条件下才能进行某试验,若具体试验分别在 30℃、60℃ 和 90℃ 条件下进行,则称“温度”为该项试验中的试验因素,“30℃、60℃ 和 90℃”分别为其三个水平。还有一种特殊的因素,其水平由事物对立的两个方面组成。如“A 药用否(即用 A 药与不用 A 药)”、“是否给予特定剂量的某种射线的照射(即用特定剂量某种射线照射与不照射)”。

所有对试验结果有影响的因素统称为影响因素,包括试验因素与非试验因素。试验因素是研究者希望通过试验着重考察的,按其性质可分为:物理的(如针刺、射线、理疗等)、化学的(如药物、毒物等)、生物的(如细菌、病毒等)以及社会和心理的因素等。而那些与试验因素同时存在,能使受试对象产生效应的其他因素被称为非试验因素,一般是受试对象自身具有的、可影响其发展过程的某些属性(如性别、年龄、血型等)或状态(如病型、病情、患病时间、生理或心理所处的时期等)。因考察的效应指标不同,有些非试验因素对试验结果的影响是不容忽视的(被称为重要的非试验因素),必须在试验设计阶段或在统计分析阶段加以控制;而有些非试验因素是可以忽略不计的(被称为次要的非试验因素)。值得注意的是,考察哪些试验因素取决于研究者的研究目的和可以投入试验研究的人力、物力、时间和经费的多少;而考察哪些重要非试验因素取决于试验分组后组间均衡性的高低、研究结果适用范围的大小和所得结论可信度的高低。例如,探讨高血压性脑出血微创术对清除血肿后脑水肿患者血浆细胞间黏附分子-1 的影响,试验因素为是否采用高血压性脑出血微创术,治疗前的血浆细胞间黏附分子-1 含量可能就是一个定量的重要非试验因素。

2) 如何确定试验因素和非试验因素

试验研究目的不同,对试验的要求也不同。若在整个试验过程中影响观察结果的因素很多,就必须结合专业知识,对众多的因素做全面分析,必要时做一些预试验,区分哪些是重要的试验因素,哪些是重要的非试验因素,以便选用合适的试验设计类型妥善安排这些因素。任何试验效应都

是多种因素综合作用的结果,我们不可能也没有必要在一次或几次试验中,同时研究所能想到的一切有关因素,只能在设法严格控制某些重要非试验因素的前提下,抓住对试验结果可能有较大影响的重要试验因素加以研究。通常,将重要试验因素简称为试验因素,将重要非试验因素简称为区组因素。例如,在研究某病患者生存期的影响因素时,研究者选定的试验因素为治疗方法(综合治疗与西医治疗),考察的重要非试验因素有性别、原始状态、TNM分期等;而在研究“辟秽防感香囊预防流行性感冒的疗效及其免疫调节机制”时,研究者选定的试验因素为是否使用辟秽防感香囊(试验组与空白对照组),考察的重要非试验因素有幼儿园儿童所在的班级、性别、年龄。

以研究某种疾病的治疗效果为例,与其有关的因素有很多,如治疗方法的种类、药物剂量的大小、治疗的疗程长短等常作为试验因素;而年龄、性别、病型、病程、是否抽烟、是否饮酒、职业、在某些方面是否有家族史、患者或研究者的心理状态等常作为重要的非试验因素。如果我们选定其中的两个因素(假设各有10个水平),就有 $10^2=100$ 种试验条件;若选定4个因素,各取10个水平,那就有10000种试验条件。每种试验条件下至少要做2次独立重复试验(具体应重复几次,应有根据地估算出来),因此,一个好的试验设计,首先要求在众多试验因素中抓住主要的几个,水平数也不能定得太多;需要控制的非试验因素往往也很多,应抓住绝对不应忽视的少量的几个重要非试验因素加以控制。控制的方法有两个,其一,按重要非试验因素组合成“复合型区组因素”,按其水平将全部受试对象进行分组,再将每组中的受试对象完全随机地分配到由重要试验因素的全部水平组合所形成的全部小组中去,这种做法被称为分层随机化。例如,将全部某病患者同时按性别(男、女)、年龄(分为青、中、老)、病情(分为轻、中、重)先分为18个小组,再将每个小组中的患者随机均分为两组,分别进入试验组与对照组。此时,复合型区组因素由“性别、年龄、病情”三个重要非试验因素复合而成。其二,分组和分析数据时将试验因素与区组因素同等看待。有时不便采用分层随机化消除区组因素对观测结果的影响,有时采用了分层随机化,似乎觉得区组因素对任何试验因素全部水平组的影响完全相同,因而在对试验资料进行统计分析时,认为区组因素的影响完全被“抵消”掉了。其实不然,每个区组因素不同水平对观测结果的影响将与真正的“随机误差”混杂在一起,此时对误差的标准误差的估计是不准确的,从而导致统计分析结论的可信度明显下降。这就好比用“单因素多水平设计一元定量资料方差分析”去处理一个区组因素有统计学意义的随机区组设计一元定量资料一样,其分析结果是错误的。

3) 合理安排试验因素与区组因素

不同因素或同一因素的不同水平,就造成试验条件的多样性,包括:①单因素单水平(本质上应属于单因素两水平)设计,这是科研中较常见的一种试验设计类型,如一种新的治疗方案对某种疾病的疗效,比较时常以“常规疗法”的特定结果(如理论平均值或率)为基准,就属于这类设计;②单因素多水平设计,比较某药物不同剂量对某病的治疗效果,属于这类设计,特别是珍贵药物、毒性较大的药物或最佳新药剂量的摸索,往往需要采用这类设计;③多因素多水平设计,例如探索联合用药方案、研究中药复方等,由于生物体的复杂性,决定了多因素多水平研究成为最常用的试验设计类型。

若试验中存在重要的非试验因素,设计时应有意地控制和消除其干扰和影响。若非试验因素是定性的,如性别等,可以采用分层随机化的方法对其加以控制,常见的区组随机化和配对随机化可看作是分层随机化的实际应用。此类试验设计包括配对设计、随机区组设计、平衡不完全随机区组设计、含区组因素的析因设计等。有些非试验因素虽然是定量的,但可以通过分组将它看作定性的因素,如年龄是定量的,若按照年龄大小分为儿童、少年、青年、中年、老年,就转变成定性的因素了,也可以按上述方法安排这种定量的非试验因素。若重要的非试验因素是定量的又不便将其定性化,则应从试验开始之时,就记录每个受试对象的定量影响因素的数值大小,将来进行数据

处理时采用“协方差分析”等技术将其影响降至最低程度。例如,研究不同饲料对动物体重的影响,饲料种类是试验因素,而动物进食量是一个很重要的非试验因素,在试验设计阶段难以严格控制,可以在试验中记录动物的每餐进食量,最后可求出每只动物在整个试验时期内的平均每餐进食量,将来进行数据处理时采用相应设计定量资料的协方差分析方法,来排除进食量对试验结果的影响。像“进食量”这样的重要的定量非试验因素,只有在试验之前就考虑到并在试验过程中认真记录下来,否则,等试验结束了再来弥补,就是绝对不可能的了!

4) 注意因素之间的交互作用

多种因素之间往往存在交互作用。交互作用是在原效应的基础上产生增强效应或减弱效应的作用,最易于理解的是两药共用可发生药效的协同作用或拮抗作用。事实上,不仅两种药物同时使用可能会产生交互作用,非药物的试验因素之间也常可产生交互作用。所谓交互作用,就是一个因素不同水平对观测结果的影响情况会随着另一个因素水平的改变而改变。因此在试验设计时除了要考虑各因素单独施加于受试对象的效应,还应特别关注因素配合施加于受试对象后产生的效应。

在试验设计阶段如何关注交互作用问题呢?选择试验设计类型时需要清楚以下几点:其一,有些试验设计类型绝对不允许因素之间存在交互作用,如无重复试验的双因素或多因素设计、无重复试验随机区组设计、无重复试验平衡不完全随机区组设计、无重复试验拉丁方设计、无重复试验交叉设计、系统分组(或嵌套)设计;其二,有些试验设计类型可以事先有意识地去安排拟考察某些试验因素之间的部分交互作用,如某些正交设计(在设计时将需要考察交互作用的那些列空着)、裂区设计、重复测量设计、无重复试验的含区组因素的析因设计、某些分式析因设计、反应曲面设计;其三,有些试验设计类型事先无法安排交互作用的“位置”,只能在使用多重回归分析处理数据过程中采取探索性方法了解某些交互作用是否具有统计学意义,如均匀设计;其四,析因设计可以考察试验因素之间的全部各级交互作用。

5) 试验因素要标准化

试验因素标准化就是要保证试验因素的定义、水平取值和施加方法等在整个试验过程中始终如一,保持不变,包括处理因素的施加方法、强度、频率和持续时间等。例如,试验因素是药品,除应确定药品的名称、性质、成分、作用及用法外,还应明确生产厂家、药品批号、出厂日期及保存方法等;若试验因素是某种手术方法,操作者的熟练程度应当自始至终保持恒定,否则将会影响结果的稳定性。同样,其他因素也必须遵循标准化原则,否则那些非试验因素或多或少会影响试验结果,导致结论出现较大偏差。因此,在设计时应制定或摸索出标准化的具体措施和方法,使试验因素真正达到标准化,并尽可能使非试验因素始终处于可控状态。

6) 实际工作者在确定试验因素方面常犯的错误

(1) 喜欢用“组别”这个词作为全部试验分组的总称,无论试验中涉及多少个组,他们总把“组别”、“处理”或“方案”等理解为“一个试验因素”,很自然地将其下的各组理解为该因素的各个水平。“组别、处理、方案”是导致人们误用统计分析方法处理给定资料的诱惑力最强的“迷魂药”。

(2) 毫无根据地选择很多试验因素,使试验研究变得十分复杂,往往因为试验条件、人力物力、财力、技术水平和时间都达不到起码的要求,使试验研究半途而废;或因安排不当,组间夹杂着某些混杂因素的影响,导致结论可信度低。

(3) 由于研究者试验设计水平贫乏,不会科学地利用多因素试验设计技术,只能将其他试验因素视而不见或固定在各自的特定水平上,每次只让自己关心的某个试验因素取不同的水平(此类错误被称为“用多次单因素设计取代相应的某种多因素设计”),当试验因素之间不独立时,基于这样的试验设计方案指导所获得的结果,即使统计分析方法选择正确,也难免得出错误的结论。

4. 把握观测指标的要领

1) 何为观测指标

一般而言,观测指标可分为两大类:其一,反映受试对象的自然属性、生活习惯或基本状况的指标,如性别、年龄、血型、是否吸烟、是否饮酒、身高、体重、血压等;其二,不同药物或治疗方法对生物体产生某种效应,体现这种效应大小的项目被称为效应指标。在临床研究中,人们所关注的指标通常指效应指标,习惯上称为观测指标。

观测指标是反映试验因素作用到受试对象身上后所产生的作用强弱的一种具体体现,它好比是一把试验因素作用效应大小的“尺子”,必须结合试验因素的性质和特点、仪器、试剂和技术水平等多方面综合考虑,找出“特异性强、灵敏度高、准确、可靠”的观测指标,以“客观指标”为主,以“半客观和主观指标”为辅。如果指标选择不当,未能准确反映试验因素的作用效果,获得的研究结果就缺乏科学性,因此,选择好的指标是关系研究成败的重要环节。例如,在考察一种降压药的治疗效果时,用药前后血压值之差值就是研究者关心的“观测指标”;在反映肝脏功能状况时,常选用谷丙转氨酶、谷草转氨酶、肝大指数、硫酸锌浊度、胎甲球和乙肝表面抗原等指标;在反映人脑老化程度时,研究者常选用年龄、图片记忆水平、数字广度记忆水平、图形顺序记忆水平、心算位数、心算时间、规定时间内穿孔数、步距、步行时双下肢夹角和步速等指标;在研究最大呼气流速-容量曲线时,研究者常选用用力肺活量、第一秒用力肺活量、最大呼气中期流速、用力呼气后期流速、呼出50%肺容量时最大流速和呼出75%肺容量时最大流速;在研究某些癌症患者经治疗后常观察生存期和无病生存期作为主要疗效评价指标;在研究某中药预防流行性感冒的疗效时,研究者选定患病率、平均病程,以及症状、体征评分为主要观察指标。

2) 观测指标的性质与特点

从资料性质角度看,观测指标可分为定量的(进一步可细分为计量的与计数的)、定性的(进一步可细分为二值的、多值名义的和多值有序的);从资料的准确程度角度看,观测指标可分为客观的(指标的每个取值都是准确可靠的)、半客观的(观测结果中夹杂着观测者个人的意志,如从X线片上读取数据,此类指标称为影像学指标)和主观的(指标的取值完全由测定者主观意愿确定的);从测定方法与测定结果之间关系角度看,观测指标可分为特异性强的指标(样品中不含有某种物质或成分时,测定的结果一定为0或阴性)与灵敏度高的指标(即使样品中含有很微量的某种物质或成分时,检测方法都能将其具体含量检测出来);从指标被选用的时间先后角度看,指标可分为老指标与新指标。进行类似的科研时,以前采用的某些指标,现在看来不够准确,随着仪器设备、试剂和测定技术水平的不断现代化,人们又会提出许多新指标取代老指标,这就是所谓指标的先进性,它是指要选择高、精、尖、新的指标。研究课题的先进性一般由指标的先进性体现出来,先进指标更能深入地反映所研究的问题,揭示事物的本质。

3) 如何选择观测指标

由于指标是用来反映试验因素作用于受试对象后所产生的试验效应大小的尺子,故指标的选择必须注意它的针对性,即选用的指标必须与所研究的目的具有本质性联系,且能确切反映试验因素的真实效应,否则,不可能作出正确的判断。例如,在评价一种治疗骨折药物疗效的临床试验研究中,研究者确定的主要疗效指标为“骨密度”、“骨钙素”、“胫骨缺损处周长”等。显然,选择这些指标来评价药物对骨折的治疗效果既不敏感又不特异,失去了“针对性”,这些指标用来反映患者“骨质疏松状况”比较合适。那么,哪些指标用来反映骨折愈合效果合适呢?根据基本常识和骨科专业知识得知,应选择“骨形态”、“X线片”、“病理检查结果”和“功能恢复情况”等指标为宜。

除了要注意指标的针对性和先进性外,选取指标时还应注意的问题有:①指标的合理性,所谓“合理”,就是指所选指标能真实客观地反映出干预措施的临床效应,在专业上能得到合理的解释。②指标的客观性,在临床试验中应该尽量选择客观性指标,即通过测试仪器或工具获得观测结果的指标,应尽量避免选用由研究者主观判断观测结果或根据受试对象主动获取观测结果的主观性指标。若不得不使用某些主观指标时,最好制定出较为严格的评判标准、对资料收集者进行严格技术培训,使主观指标趋向于客观化。③指标的灵敏性和特异性,灵敏性是指所选用的指标对干预措施反应的灵敏程度,指标的特异性即检测结果的专一性,以便把假阳性结果控制在最低水平,最好是选用灵敏度高、特异性又好的指标。④指标的精确性,包括指标的精密性(即多次测定结果与其算术平均值都非常接近,简称重现性好)与准确度(多次测定结果与其真值都非常接近,简称准确度高)的两重含义。⑤指标的经济性,检测成本应在受试对象所能承受的范围之内。⑥指标的可操作性,指标取值的测定通常需要相应的仪器设备或试剂,若选定的指标过新或对灵敏度要求过高,目前的仪器设备或试剂还达不到要求,就属于指标的可操作性差。

4) 指标的观察

选好反映试验效应的指标以后,还要规定指标观察的常规方法,如观察方法、标准、时间、记录方法及记录格式等。指标的观察或测量应避免偏性,否则会影响结果的准确度。有些指标属于整体表现(气色、精神状态、营养状态等)、主观症状(疼痛、抑郁、焦虑等)、动物行为学变化(活跃与否、异常行为等)及某些形态学变化(如器官外观、动物毛色、红肿、溃烂等),这些指标的判断带有主观性,研究者的心理常偏向于阳性结果,医生常偏向于新疗法组。观测指标的时间点的选定也是至关重要的。例如,若把治疗普通感冒的结果观测定在服药后两周未,那就不够妥当了。因为即使不给予药物治疗,普通感冒在两周后一般也会自愈;有一些药物的疗效并不能在某个确定的时间点上完全表现出来,需要多观测几个时间点(时间点的间隔多长也应依据专业知识来定),方可比较全面地反映此类药物的真实疗效。

减少偏性的措施之一是采用盲法设计:如果试验效应的观察带有偏性,则会影响观测结果的准确性。为了消除或最大限度地减少这种偏性,在试验设计时常采用盲法(Blind Method)。单盲法(Single Blind Method):仅受试对象不知道自己被分在哪一组;双盲法(Double Blind Method):受试对象和试验执行者均不知道受试对象被分在哪一组;三盲法(Triple Blind Method):受试对象、试验执行者和统计分析人员三者均不知道受试对象被分在哪一组。双盲法在临床试验中应用非常广泛,三盲法更加科学严谨,建议尽可能多多采用。

减少偏性的措施二是对指标进行标准化:指标的标准化是指指标的采集方法、部位、保存、运输、测定时间、测定方法、测定条件等方面的固定化。不进行标准化,这些与测定结果准确与否有关的非试验因素就不可避免地成为混杂因素,使指标的测定不稳定,必然影响研究结果的准确性。

例如,为探讨全脑血管造影术对患者情绪的影响及心理干预对其的作用,原研究者采用心理因素调查表对81例行全脑血管造影术的患者进行调查,并用汉密尔顿焦虑量表、汉密尔顿抑郁量表、焦虑自评量表和抑郁自评量表比较心理干预前后患者的焦虑、抑郁情况及不同疾病患者的焦虑、抑郁情况。本试验的观测指标是患者的焦虑、抑郁情况。病人的心理状况是主观指标,原文作者运用汉密尔顿焦虑量表、汉密尔顿抑郁量表、焦虑自评量表和抑郁自评量表进行评价,将其客观化,在一定程度上减少了偏性,从而使结论比较可信。然而,以各量表所得总分作为主要评价指标,其科学性与合理性尚存疑虑。

5) 主要与次要指标

在临床科研(特别是心理测量、中医中药研究等)中,就如何选择观测指标这一环节来看,有一

个普遍存在且不容忽视的问题：研究者要么罗列出几十项观测指标，且不告知哪些是主要的有效性指标、哪些是次要的有效性指标；要么运用几十项、甚至上百项内容的“测量量表”，并采用量表中多项得分之和作为主要评价指标的观测值。第一种做法的不妥之处在于：在众多指标中分不清主次，导致“眉毛胡须一把抓”和“出现互相矛盾的分析结果”等混乱局面和无法做出肯定结论的尴尬结局。第二种做法的不妥之处在于：从表面上看，用量表的总分反映每位受试者的试验效应似乎是合情合理的，其实不然，因为量表中有些项目是高优指标（即指标取值越大越好），还有些项目是低优指标（即指标取值越小越好）；另外，众多指标彼此之间的重要性程度不同，某两项指标同样得8分，但反映患者病情的实际改善程度可能相差很多。由于上述两方面原因，对量表中的几十项甚至上百项内容所得分值简单求和作为主要评价指标的取值，其科学性与合理性难免受到质疑。

比较科学且具有可操作性的做法是：结合基本常识和专业知识，并汇集众多同行专家的意见和建议，从众多的疗效指标中筛选出少数（一二个）主要指标，同样，从众多的安全性指标中筛选出少数（一二个）主要指标，其他指标一律视为次要指标。若实在无法找出少量的主要指标，可找出5到10个相对重要的指标，先对它们进行“趋同化处理”，即使它们都转变成“高优指标”或“低优指标”，再采用综合评价方法将它们合并成一个复合型的“主要指标”。

确定一个研究项目中的主要指标至少有以下三点好处：其一，重点突出；其二，便于试验前估算样本含量；其三，可以有效避免统计分析结果出现自相矛盾的现象。

11.3.3 试验设计四原则及把握的要领

1. 把握随机原则的要领

1) 随机原则的概念与作用

何为随机原则呢？就是在选取样本时，应确保总体中任何一个个体都有同等的机会被抽取进入样本；在分配样本时，应确保样本中任何一个个体都有同等的机会被分入任何一个组中去。这就是严格意义上的随机化原则。

随机原则的作用是什么？其作用就是消除人为因素和心理因素对抽样或分组造成的偏性影响，使样本具有极好的代表性，使各组受试对象在重要的非试验因素方面具有极好的均衡性，提高试验资料的可比性。

随机化应该体现在如下几个方面：

(1) 抽样随机：每一个符合条件的试验对象参加试验的机会相同，即总体中每个个体有相同的会被抽取进入样本之中。

(2) 分组随机：每个试验对象被分配到不同组（通常为对照组、不同处理组）中去的机会相同。其中，第(1)条保证所得到的样本具有代表性，使试验结论具有普遍意义；第(2)条保证各组间试验对象尽可能均衡一致，以提高各组间的可比性。

某研究者闭着眼睛从装有很多小鼠的笼子里随意地抓小鼠。将先抓到的10只小鼠分到安慰剂组，将第二批抓到的10只小鼠分到甲药组，将最后抓到的10只小鼠分到乙药组，这样做科学吗？严格地说是科学的！这实际上是误认为“随便”就是“随机”。事实上，即使闭着眼睛盲目地抓小鼠，相对来说，先被抓到的小鼠活动能力和健康等状况较差，而后被抓到的小鼠状况可能相反。这样一来，3个组之间除了药物不同以外，受试对象的个体差异仍相差悬殊，若试验结果间有差别，就没有充分的理由认为是因为它们服了不同的药物所引起的。也就是说，在进行试验分组时必须采取“随机化”措施，以消除“重要的非试验因素”对观察结果的干扰和影响。

那么,完全随机化的效果是否永远最好?不一定,关键取决于样本含量的大小。若样本含量很大,完全随机化的效果应当是比较理想的;若被随机化的样本比较小,则完全随机化产生的各组受试对象之间在很多重要的非试验因素方面可能参差不齐。此时,采取分层随机化法效果会更好一些。即先按某些重要的非试验因素(取决于他们是否对观测指标有较大影响)将受试对象分组,然后对每个小组(在所考虑的重要非试验因素上条件一致)中的受试对象进行完全随机化,使他们被均分到各试验组中去,这样才能确保各试验组中的受试对象之间在所考虑的重要非试验因素方面保持均衡一致。例如,现有 24 只小鼠,其中 16 只雌性,8 只雄性,若采用完全随机化方式将它们均分成 2 组,则有可能某一组中的 12 只小鼠全是雌性的,另一组中有 4 只雌性、8 只雄性小鼠。若果真出现这样的分组结果,当小鼠性别对观测结果又有较大影响时,“性别”这个重要的非试验因素就严重地影响了试验因素两个水平(处理组与对照组)对观测结果效应大小的正确评价,甚至易于得出歪曲事实的结论来。此时,宜采用“分层随机化”,就本例而言,先用完全随机化法将 16 只雌性小鼠随机均分成两组,试验组和对照组各 8 只;再用完全随机化法将 8 只雄性小鼠随机均分成两组,试验组和对照组各 4 只。这样,试验组和对照组都各有 12 只小鼠,其中雌性和雄性小鼠分别为 8 只与 4 只,即“性别”这个重要的非试验因素在试验组和对照组中所施加的影响是均衡的。

2) 实现随机化的一般方法

实现随机化的方法有多种,可以采用“抽签”法、查“随机数字表”、“随机排列表”、查“用计算机产生的伪随机数字表”或直接由计算机程序来实现。

3) 完全随机化的一般步骤

第一步,将受试对象编号,将编号按顺序写成一排。

第二步,事先规定分组的规则。如分两组时,可规定遇到随机数字为偶数时将对应的受试对象分入试验组、遇到随机数字为奇数时将对应的受试对象分入对照组(反过来规定也可以);再如分三组时,可事先规定,凡随机数字除以 3 得余数为 0 者分入第一组,余数为 1 者分入第二组,余数为 2 者分入第三组。当然,也可以规定其他的分组规则,但规则必须事先确定下来,一旦确定不应随意改动。

第三步,从上述三种随机数字表中任意指定的位置开始向后(或向前)抄录随机数字,依次写在各编号之下,注意:舍弃不符合要求的随机数字(如随机数字超过了编号所对应的数字)。

第四步,根据抄录的随机数字按事先确定的分组规则分组。当各组样本含量不等时最好再用随机的方法进行调整,尽可能使各组的样本含量相等或接近相等(从统计计算角度看,各组样本含量相等时误差较小)。若采用的是“随机排列表”,一般来说,分组的结果会使各组样本含量相等,这样麻烦事就少多了。当然,现在可以运用计算机程序直接实现随机化分组,使随机化的过程大为简化,非常方便。

4) 动态随机化法

在临床试验研究中,常习惯将患者分为试验组与对照组,患者来医院就诊,若将先来就诊的分入试验组,后来就诊的分入对照组,是不正确的。因为患者来医院就诊,在患病严重程度、患病时间等重要非试验因素方面不一定是均衡的,可能带有某种程度上的“聚集性”,即某一段时间上来的患者多数为重病患者,另一段时间上来的多数为轻病患者;若规定第一个来就诊者分入试验组,第二个来就诊者分入对照组,这样交叉进行分组,最后也可能两组患者在某些重要非试验因素方面相差悬殊,组间缺乏均衡性。有一种简便易行的随机化分组方法叫“按不平衡指数最小的分配原则”分组。

2. 把握对照原则的要领

1) 何为对照原则

所谓对照,是指在试验研究中,要想得出什么处理或治疗更佳,必须交代“与谁比较”,比较至少要在两个事物之间进行,在统计学上通常是两个组之间进行,即试验组与对照组。设立对照实际上就是寻找一个参照物或对比的基础。对照原则的作用是提高试验的鉴别能力,增强结论的说服力。

需要强调一下,对照组的概念还必须予以拓展,在多因素的试验设计中,为了提高同一个因素不同水平之间的可比性和更准确地考察因素间的交互作用,往往要求试验各组都有与之对应的对照组,否则,如果对照设置不全就不能把某个因素的全部水平组进行一起分析。因此,不要以为在任何试验研究中设立了1、2个对照组,就一定能达到研究目的。

2) 为什么要设置对照组

对照原则的作用是提高试验的鉴别能力,将处理因素给受试者带来的效应(症状、体征或指标的改变)与其他因素(如疾病的自然进展、观察者或患者的期望、其他治疗措施等非处理因素)造成的效应区分开来。出自对照组的试验效应可以告诉人们,在没有接受研究者关注的那种特定处理的条件下,患者发生了什么情况。“王婆卖瓜自卖自夸”等俗语都清楚地说明一个道理:缺乏对照的研究是没有说服力的。

例如,为探讨某种新药治疗急性脑梗死的临床疗效,若只对一组病人给予该药物治疗,评价用药后的疗效,虽然有时或许也能观察到患者病情有所好转,但因为没有与其他现有的药物作比较,不能排除是其他非处理因素产生的效应还是此新药产生的疗效,就很难得出这种药物疗效优劣的结论。如果研究者将纳入研究的足够多的患者随机分为两组(两组患者在病情、年龄、性别构成等方面非常均衡),一组给予试验药物,另一组给予现有的常规治疗,则可以鉴别出该试验药物的真实效应。

3) 对照设置的种类

对照有很多种类,可根据研究目的和内容加以选择。

(1)安慰剂对照:安慰剂是指外观与受试药物相同,且无药理活性的物质。在临床研究中安慰剂对经常用来代替受试药物,以排除精神心理等非药物因素的影响。

(2)空白对照:即对照组不接受任何处理因素。空白对照简单易行,但在以人为对象的研究中容易引起对照组和试验组在心理上的差异,从而影响试验结果的可靠性。

(3)试验对照:对照组不施加处理因素,但施加某种与处理因素有关的基础处理。由于试验结果往往会受到处理因素和其他非处理因素影响,设立试验对照可以用来排除某种特定的重要非处理因素的干扰。例如,为了评价一种中草药烟熏剂对空气中某些细菌的杀菌效果,需要采用无药物性能的柴草点燃后产生的烟雾作为试验对照,以抵消“烟雾”本身对结果造成的影响。

(4)自身对照:对照与试验在同一受试对象身上进行。但应注意试验措施在不同时间实施时,时间间隔不应过长。若从每个个体身上取出样品,均分成两份,分别施加不同的处理,这种自身对照的结果准确度高。

(5)标准对照:用现有标准方法、常规方法、标准值或参考值作为对照。试验研究中常用于某种新方法是否可代替传统方法。

(6)历史对照:历史对照是将研究者以往的研究结果或文献上他人的研究结果与本次研究结果作对照。这种对照通常意味着试验组与对照组在受试对象、环境、条件、研究者、测定方法和评价方法等很多方面不一致,可比性差,说服力不强,应尽量少用。

4) 只设一个对照组是否总是合理的

试验设计中只设置一个对照组是否总是合理的？应视具体情况而定。若研究者的试验中仅涉及一个试验因素，一个对照组已足够。例如，研究一种药物的疗效，只要设置一个空白对照、标准对照或安慰剂对照。若研究者的试验涉及多个因素，为了能正确评价因素间的交互作用效应大小，往往要求每个试验因素都有相对应的对照组，否则，如果对照设置不全，某些试验因素就有可能成为其他试验因素的“混杂因素”了，从而破坏了那些试验因素组间的均衡性。例如，在某项研究中，观察两种药物应用与否的作用，研究者需设立 4 个组，方可全面考察两个因素单独作用的效应及它们之间的交互作用效应大小。有时在同一项试验研究（如多因素非平衡组合试验）中需要同时运用多种对照形式，如安慰剂对照和试验对照同时并存，有时甚至还会出现阴性对照、阳性对照等对照形式。因此，并非在任何试验研究中设立 1 个或 2 个对照组，就一定能达到研究目的。

5) 设置对照的原则

设立对照组，应满足均衡性的原则。通过随机化、盲法等手段，使试验组和对照组除处理因素不同外，其他重要的、可控制的、对试验结果有影响的非处理因素应尽量一致。例如，在动物试验中，动物的来源、种属、性别、体重、健康状况应尽量相近。在以人为受试对象时，受试者的年龄、性别、病情、种族等基线资料应当相同或相近。

在临床试验中，对照组还应遵循专设、同步的原则。否则，将失去对照的意义。所谓专设，是指在临床试验中，将合格的受试者分出部分受试者作为对照，在试验结束时比较两组的处理效应才能达到对照组所起的“比较鉴别”作用。一般情况下，不宜只观察试验组的结果，将该结果与研究本人或他人过去的研究结果进行比较。所谓同步，就是要求设立平行对照组，从与试验组相同的人群中选择对照，同时按照各自规定的方法进行治疗。

对照形式应与研究目的和内容相符。例如，在药物临床试验中，尽量选择安慰剂对照、标准对照或阳性药对照，而不选择空白对照，只有当安慰剂盲法试验无法执行，或执行起来非常困难，又没有合适的标准对照或阳性药对照，且不会严重违背伦理道德时，才选择空白对照。

3. 把握重复原则的要领

1) 重复原则与样本含量

试验设计中的重复原则是指在特定的条件下应做足够多次数的独立重复试验，以便使随机变量的变化规律性能充分地显露出来。而人们习惯把整个试验研究中所用到的受试对象的个数称为样本含量。事实上，应该把每种特定条件下用到的独立受试对象的个数称为样本含量更贴切、更有实际意义。因为有时一个试验研究项目仅把受试对象分为两组，但有时需要把受试对象分为 12 组、24 组、48 组，有时甚至会有更多组。

值得一提的是，在生物医学和临床研究中，“重复”还有其他两层含义：其一，重复取样，即从每位受试者身上获得一个样品，将其均分成若干份，在同一时间点上对其进行逐一观测，其目的是看各标本中某定量观测指标值的分布是否均匀或检测方法是否具有重现性；其二，重复测量，即在不同时间点或不同部位，对同一位受试者或取自其身上的样品进行观测，其目的是看定量指标随时间推移的动态变化情况或部位改变条件下定量指标的分布情况。

2) 样本含量估计和检验效能的含义

正确估计样本含量体现了统计研究设计中的重复原则，可以降低研究中的抽样误差，同时足够的样本量也是试验研究中保证组间均衡性的基础。若样本含量过小，观察指标的平均值不稳定，意

味着抽样误差大,推论总体的精密性与准确性都比较差,统计检验的效能低,实际存在的差别不易真实地显露出来;样本含量过大,会增加实际工作的困难,浪费人力、物力、财力和时间,虽然减少了抽样误差,但由于过分追求数量,可能引入更多的混杂因素或因工作粗枝大叶导致科研资料不准确,对研究结果造成不良影响。

检验效能由第二类错误概率 β 的大小决定,等于 $1-\beta$ 。其意义是,当所研究的两个总体确有差别时,按检验水准 α 能够发现它的概率。如果 $1-\beta=0.9$,则意味着当 H_0 不成立时,理论上在每100次抽样中,在检验水准为 α 时,平均有90次能拒绝 H_0 。

当假设检验出现“阴性”结果($P>0.05$)时,有必要复核样本含量和检验效能是否偏低,以便正确分析假设检验“阴性”结论的正确性。若检验效能偏低,说明试验的样本量不够,应进一步增大样本量,继续进行试验;若检验效能足够大,可下“接受 H_0 ”的阴性结论,尽管此时仍有可能犯第二类错误,但犯第二类错误的概率 β 在可接受的范围之内。

4. 把握均衡原则的要领

1) 何为均衡原则

进行一项研究,一般都会涉及两组或两组以上患者的药物疗效或治疗措施的效果之间的对比问题,若各组患者之间在很多方面几乎都一致,其区别仅仅在于药物的种类不同或同一种药物的剂量不同或手术的类型不同,则研究结果就具有较高的说服力。进行科研设计时要求各组患者之间在以下诸多方面应尽可能一致,这就是“均衡”的真实含义,即病情构成(即轻、中、重的比例),平均病程及病程分布,平均年龄及年龄分布,性别构成(男女比例)等任何可能影响疗效评价的方面,它们在统计学上常被称为重要的非试验因素。

均衡是对一个试验因素的各个水平而言的。在一项试验研究中,若同时需要考察的试验因素为两个或两个以上时,谈论“均衡”仍应限定在每个试验因素的各个水平之间。例如,在考察三种药物的疗效时,三个药物组中的患者若用了不同的剂量,则三个药物组在“剂量”这一重要的非试验因素上是不均衡的。

均衡原则就是要求同一个试验因素各水平组之间除了所考察的因素取不同水平外,在其他试验因素和一切非处理因素方面都应达到均衡一致。例如,某研究中,一组为正常人,测定其血压值,另一组是高血压患者,测定治疗前、治疗后的血压值,很多人将这三组定量数据视为来自“单因素三水平设计的定量资料”,这显然是不妥的!为什么?请读者自己独立思考后作出解答。

2) 均衡原则的作用

确保试验因素各水平组间不受其他试验因素或重要的非试验因素的不平衡所带来的干扰和影响,以便使所考察的试验因素在不同水平条件下对观测结果的影响真实地显露出来。

3) 一个违反均衡原则的经典案例

为了弄清经常吃含赖氨酸的面包是否对少年儿童生长发育有帮助,有人设计了如下的试验,即对住校的某班级学生,随机让一半学生每天上午和下午课间各吃一个含赖氨酸的面包,另一半学生课间不吃面包,一年后比较这两组学生的身体发育情况,从而得出“赖氨酸”有利于少年儿童生长发育的“荒唐结论”!因为课间学生肚子会有些饿,即使吃一个不含赖氨酸的面包也会有助于少年儿童生长发育,这两组学生在重要的非处理因素(即是否吃面包)方面是不均衡的,故其结论是不可信的。

只有随机地决定一半学生吃不含赖氨酸的面包,另一半学生吃含赖氨酸的面包(每个面包的大小和其他营养成分应相同),其他(如锻炼身体情况、作息时间等)各方面两组学生应尽可能一致,一年结束时再比较,才有可能说明“赖氨酸”的价值有多大。

均衡性的道理很简单,即对比组之间除了所考察的因素取不同水平外,其他一切方面都应尽可

能一致,但做到不容易。研究设计者要有很强的责任感,要清醒地认识到“违背均衡原则”将会得出错误结论。

4) 提高试验设计均衡性的方略

提高试验设计均衡性的方略在于研究设计者自身、同行专家和统计学工作者三方面的高度重视和深思熟虑。

第一,研究设计者必须有高度的责任感、丰富的专业知识和较扎实的统计学基础,以便使所制定出来的试验设计方案初稿具有较高的质量;第二,制订出试验设计方案后,应多请几位同行专家多补充和修改意见,使其在专业和统计学方面有较大程度的改进;第三,最好再请一两位有丰富试验设计经验的统计学工作者把关,使试验设计方案得到进一步完善,尤其是在三要素、四原则和设计类型方面,特别是其中的对照与均衡原则方面,统计学工作者往往可以提出很多富有建设性的建议。

11.3.4 试验设计类型及把握的要领

1. 何为试验设计类型

试验设计类型就是因素及其水平所决定的一种结构。这种结构不单纯取决于“形式”,还取决于“内容”。例如,在某项试验研究中涉及三个试验因素,设各试验因素都有 2 个水平,试验因素的全部水平组合共有 8 种,在这 8 种水平组合下均做了 3 次独立重复试验。现在若问:这个试验研究所对应的试验设计类型叫什么?在没有其他附加条件时,是无法判定的,因为前面所讲的这些信息实际上仅呈现了该试验设计类型的“形式”,缺少进一步用于判定试验设计类型的“内容”。若“内容”属于下面三种情形之一,就对应着三种特定的试验设计类型。

其一,因素同时施加,且各因素对观测结果的影响地位平等。即每次试验时,必然涉及每个因素的一个水平,这就是因素同时施加之意;因没有充分的专业知识为依据,无法给各因素对观测结果的影响排出主次顺序,就是各因素对观测结果的影响地位平等之意。此时结合前面所具备的“形式”,再加上现在所讲的条件,所对应的试验设计类型就称为三因素析因设计。注意:通常试验因素与受试对象的分组之间无直接的关系,也就是说,受试对象可以被完全随机地分配到任何一个多因素水平组合而形成的组中。

其二,因素同时施加,但各因素对观测结果的影响有主次之分。即每次试验时,必然涉及每个因素的一个水平,这就是因素同时施加之意。因有充分的专业知识为依据,可以将全部因素对观测结果的影响排出主次顺序,此时结合前面所具备的“形式”,再加上现在所讲的条件,所对应的试验设计类型就称为三因素系统分组设计(或称为嵌套设计)。

其三,在整个试验过程中因素的施加有先后顺序之分。即在整个试验研究过程中,全部因素并非同时施加于试验,而是分期、分批进入试验。例如,有三种瘤株分别接种于三组动物,使它们长瘤子,几天后,长出了瘤子,将每一组的动物均分成三个小组,分别用三种不同的抑制瘤子的药物治疗,最后以各小组动物瘤子的体积作为观测指标来评价三种药物对三种瘤株的抑瘤效果。显然,在此项试验研究中涉及“瘤株种类”与“抑瘤药物种类”两个试验因素,前者先作用于动物,后者后作用于动物,先后之间有一定的时间间隔。此时结合前面所具备的“形式”,再加上现在所讲的条件,所对应的试验设计类型就称为两因素分割设计(或称为裂区设计)。

综上所述,试验设计类型需要根据因素及其水平的组合情况、因素作用于受试对象的时间、因素对观测结果的影响情况等方面来决定。掌握各种标准的试验设计类型,对科学地安排试验、合理选用统计分析方法处理定量资料都是极其有用的。

2. 试验设计类型的种类

1)按因素与水平被放置的方向来划分

(1)以横向与纵向两个方向放置因素及其水平的设计类型——交叉表结构的设计类型

特别是在两因素及两因素以上的试验场合下，将因素及其水平分别放置在横向与纵向两个方向上，由不同因素的不同水平组合决定的条件称为“试验点”，全部试验点分布在横向与纵向交叉处，见表 11-4。

表 11-4 3 个试验因素作用下 OD 值的测定结果

OD 值									
因 素 A 与 B		因素 C:C ₁		C ₂		C ₃		C ₄	
A ₁	B ₁	0.39	0.41	0.37	0.39	0.42	0.38	0.44	0.41
	B ₂	0.37	0.36	0.43	0.45	0.41	0.37	0.42	0.39
	B ₃	0.45	0.43	0.46	0.39	0.38	0.35	0.39	0.37
A ₂	B ₁	0.36	0.41	0.45	0.36	0.41	0.45	0.41	0.46
	B ₂	0.42	0.37	0.38	0.41	0.38	0.36	0.43	0.38
	B ₃	0.37	0.43	0.36	0.39	0.43	0.42	0.35	0.37

说明，此表反映了某试验同时涉及 A、B、C 3 个地位平等的试验因素，A 分为 2 个水平、B 分为 3 个水平，它们被安置在横向上；而 C 分为 4 个水平，它被安置在纵向上。观测指标为 OD 值，受试对象为样品，不同试验条件下均独立地重复做了 2 个样品，此为三因素析因设计。

单因素设计的因素与水平可以仅放置在横向或纵向上，而随机区组设计、平衡不完全随机区组设计、无重复试验的多因素设计、交叉设计、拉丁方设计、嵌套设计、裂区设计、重复测量设计、析因设计和含区组因素的析因设计等，都需要借助纵横两个方向来放置因素与水平。

(2)仅以纵向一个方向放置因素及其水平的设计类型——数据库结构的设计类型

当因素很多时，仍以纵横两个方向来说明不同因素不同水平组合所决定的试验点并不太方便，将每个因素及其水平改为从上到下的纵向放置，全部因素及其水平按列排放，按行看水平组合，见表 11-5。

表 11-5 利用 L₈(2⁷) 正交表安排两道工序试验的裂区设计及试验结果

试 验 号	A 1	B 2	一级误差 3	二级误差 4	C 5	D 6	E 7	镇静率 P
1	1	1	1	1	1	1	1	5/15
2	1	1	1	2	2	2	2	13/15
3	1	2	2	1	1	2	2	14/15
4	1	2	2	2	2	1	1	11/15
5	2	1	2	1	2	1	2	10/15
6	2	1	2	2	1	2	1	2/15
7	2	2	1	1	2	2	1	7/15
8	2	2	1	2	1	1	2	13/15

此为正交设计。此表中 A、B、C、D、E 的含义及其水平见表 11-6。

类似的试验设计类型还有分式析因设计、均匀设计、各种回归设计、各种最优设计、混料设计等。

2) 按拟采用的处理定量资料的统计分析方法来划分

(1) 差异性检验的设计类型。来自单因素设计、随机区组设计、平衡不完全随机区组设计、无重复试验的多因素设计、交叉设计、拉丁方设计、嵌套设计、裂区设计、重复测量设计、析因设计、含区组因素的析因设计和正交设计等的定量资料, 一般都可采用差异性检验, 通常包括 t 检验、方差分析或秩和检验。

表 11-6 研究灵芝的镇静作用时考察的试验因素及其水平

试验因素	因素各水平的具体内容	
	水平代码: 1	2
A(溶剂种类)	10 倍量 70% 乙醇渗漉	10 倍水煎煮 2 次
B(浓缩温度)	100℃ 常压浓缩	70℃ 减压浓缩
C(给药天数)	当天给药	连续给药 3 天
D(给药途径)	腹腔注射	灌胃
E(给药剂量)	0.1ml/10g 体重	0.2ml/10g 体重

(2) 回归分析的设计类型。来自分式析因设计、均匀设计、各种回归设计、各种最优设计、混料设计等的定量资料, 一般都需要选用多重回归分析方法处理。

3) 按是否可对定量资料直接进行统计分析来划分

(1) 标准的设计类型。在统计学上能给出具体设计名称的那些设计类型被称为标准的设计类型, 即前面提到的那些试验设计类型。

(2) 多因素非平衡的组合试验。有些试验涉及多个试验因素, 但研究者并没有对因素的水平进行全面组合, 所做的试验点只是全面组合中的很少一部分, 为什么仅做一小部分试验点, 没有任何数学或统计学依据(正交设计有时也仅做了全面试验的一小部分, 但试验中所考察的那些试验点在数学上满足正交性; 而均匀设计中的一小部分试验点在数学上满足均匀性)。对这样的试验安排, 可以形象地比如为: 研究者“拍着脑袋定方案、随心所欲搞科研”的产物。笔者称其为“多因素非平衡的组合试验”, 因为这种安排下的试验数据不便直接进行统计分析, 需要结合专业知识和统计学知识, 对全部组进行合理拆分, 并进行重新组合, 使每种新组合所包含的那些组在专业上具有很好的可比性, 在统计学上又是一个标准的设计类型。这种情况在生物医学和临床研究中是常见的, 参见下面的实例。

【例 11-4】 某研究者为研究 bFGF 对肌成纤维细胞的作用及其对创面愈合的影响, 制作大鼠 30% 深 II° 烫伤模型。观察的指标有 α -平滑肌肌动蛋白等定量指标。

【错误的试验设计】 原作者设计的三个组如下, 每组 10 只大鼠。

(1) 正常对照组(将大鼠背部皮肤置于 37℃ 水中 8s), 切取创面组织检测多项定量观测指标的数值;

(2) 单纯损伤组(造成 30% TBSA 深 II° 烫伤), 在伤后 3h 切取创面组织检测多项定量观测指标的数值;

(3) 在(2)的基础上, 用 bFGF 治疗。在伤前、伤后 3h、6h、1d、3d、7d、14d 切取创面组织检测多项定量观测指标的数值。

将第(3)组产生的 7 个小组数据分别与第(1)、(2)组比较, 用 t 检验进行统计分析。

这个试验设计是有严重问题的。

【对差错的辨析与释疑】 问题出在第(2)组仅观测了伤后 3h 的结果, 但指标的取值会随着时间的推移而发生变化。由于其他时间点上指标的取值未观测, 故情况并不清楚。将第(3)大组中的 7 个不同时间点上观测的结果与第(2)组一个固定时间点上结果进行比较, 仅伤后 3h 这一个时间点上的两组数据之间的比较具有可比性, 另外, 尚缺少一个“(正常 + 用药)组”。所以说原研究者的试验设计是错误的。

【正确的试验设计】 首先应考虑多次取创面组织进行检测, 其“刺激”本身是否会引起指标取

值的改变,若不会改变,正常对照组仅观测一次就可以了;若会改变,应像第(3)组一样取7个观测点;无论第(1)组如何处置,第(2)组都应像第(3)组那样取7个观测点。下面分三种情形来讨论。

情形之一:若第(1)组只需取一个时间点,则正确的试验设计如下:

(1)正常对照组(将大鼠背部皮肤置于37℃水中8s),切取创面组织检测多项定量观测指标的数值;

(2)单纯损伤组(造成30%TBSA深Ⅱ°烫伤),在伤前、伤后3h、6h、1d、3d、7d、14d切取创面组织检测多项定量观测指标的数值;

(3)在(2)的基础上,用bFGF治疗。在伤前、伤后3h、6h、1d、3d、7d、14d切取创面组织检测多项定量观测指标的数值。

此时的试验设计类型是什么呢?是“多因素非平衡组合试验”,因为第(1)组只有一个时间点,与另两组不平衡,需要进行必要的拆分。拆分的方法如下:

组合1:正常对照组、第(2)组伤前、第(3)组伤前;

组合2:第(2)组全部时间点;

组合3:第(2)组和第(3)组全部时间点。

说明:组合1属于“单因素三水平设计”,本质上属于三组大鼠“基线”的比较;组合2属于“具有一个重复测量的单因素设计”,可以考察伤后各时间点上指标取值相对于伤前指标取值的变化情况,其设计格式见表11-7;组合3属于“具有一个重复测量的两因素设计”,可以考察“是否使用bFGF”、“重复观测时间点”以及这两个因素之间的交互作用对观测指标的影响大小,设计格式见表11-8。当然,组合2的存在意义不大,因为组合3也可实现组合2的分析目的,此处给出组合2意在介绍“具有一个重复测量的单因素设计”这一设计类型。

表 11-7 单纯损伤组伤前和伤后6个时间点上某定量指标的测定结果

动物 编号	观 测 指 标						
	伤前	* :3h	6h	1d	3d	7d	14d
1	X	X	X	X	X	X	X
...
10	X	X	X	X	X	X	X

注:表中各“X”代表每只动物各时间点上某定量观测指标的取值;* :代表“伤后观测时间”。

表 11-8 单纯损伤组与烫伤后用药组伤前和伤后6个时间点上某定量指标的测定结果

是否 用药	观测指标($\bar{x} \pm s$)						
	伤前	* :3h	6h	1d	3d	7d	14d
不用	X	X	X	X	X	X	X
用	X	X	X	X	X	X	X

注:表中各“X”代表每只动物各时间点上某定量观测指标的($\bar{x} \pm s$),不用药组即第(2)组,用药组即第(3)组;* :代表“伤后观测时间”。

情形之二:若第(1)组与另两组测定的时间点一样多,则正确的试验设计如下:

(1)正常对照组(将大鼠背部皮肤置于37℃水中8s),在伤前、伤后3h、6h、1d、3d、7d、14d切取创面组织检测多项定量观测指标的数值;

(2)单纯损伤组(造成30%TBSA深Ⅱ°烫伤),在伤前、伤后3h、6h、1d、3d、7d、14d切取创面组织检测多项定量观测指标的数值;

(3)在(2)的基础上,用bFGF治疗。在伤前、伤后3h、6h、1d、3d、7d、14d切取创面组织检测多项定量观测指标的数值。

此时的试验设计类型是什么呢?仍是“多因素非平衡组合试验”,因为“正常对照组”、“单纯损伤组”与“(损伤+用药)组”是“是否损伤”和“是否用药”两个因素的4种水平组合中的3种,属于对照不全,需要拆分。拆分的方法如下:

组合1:第(1)组与第(2)组;

组合2:第(2)组与第(3)组。

说明:上述两种组合所对应的试验设计名称是相同的,即都是“具有一个重复测量的两因素设计”,其试验设计的表格形式与表11-8相同。

情形三:在情形二的基础上,再增加一个大组,即“正常对照+用药”,也在相应的7个时间点上观测指标的取值。此时,对应的试验设计类型称为“具有一个重复测量的三因素设计”,其设计格式见表11-9。

表 11-9 烫伤与否和用药与否在伤前和伤后6个时间点上某定量指标的测定结果

烫伤 与否	用药 与否	观测指标 ($\bar{x} \pm s$)						
		伤前	*:3h	6h	1d	3d	7d	14d
否	否	X	X	X	X	X	X	X
	用	X	X	X	X	X	X	X
用	否	X	X	X	X	X	X	X
	用	X	X	X	X	X	X	X

注:表中各“X”代表每只动物各时间点上某定量观测指标的($\bar{x} \pm s$),未烫伤不用药组即第(1)组,烫伤不用药组即第(2)组,烫伤用药组即第(3)组;* :代表“伤后观测时间”。

11.3.5 如何进行试验过程中的质量控制

1. 进行质量控制的必要性

无论试验设计方案制定得多么科学严谨,它毕竟是纸上的东西,要把纸上写的东西落到实处,需要通过实施者采取具体的行动,需要受试对象(特别是人)的全力配合,需要环境和条件处于设计者所希望的状态,然而,在较长时间的试验过程中,前述提及的诸多方面都或多或少会发生一些变化,一旦那些变化超出了“允许”的范围,其试验结果就会偏离其真实值较远,结论的可信度就必然会大打折扣。因此,对试验过程中出现的可能会影响结果准确度的任何问题,都应能及时发现,并能在尽可能短的时间内消除其影响。因此,在试验过程中,进行质量控制是非常必要的。

2. 进行质量控制的环节与措施

需要进行质量控制的环节包括:控制研究者的责任心、心理状态和技术水平对试验结果可能造成的不利影响;控制受试者的心理状态和不依从性对试验结果可能造成的不利影响;控制环境的变化(如温度、湿度、光线等)对试验结果可能造成的不利影响;控制试验条件的改变(如仪器的状态发生改变、试剂过期、观测指标的单位不同、测定的方法不同等)对试验结果可能造成的不利影响。

对试验过程中的质量进行控制的措施有:对不同单位的多名研究者应进行统一的培训,内容包括增强责任心教育、调整好心理状态(在言行上不应有个人的倾向性或偏好)、接受正规的技术培训,并要求每一位研究者应自觉遵守统一的操作规程。对于动物或样品类的受试对象而言,应当选用与研究目的和试验因素相对应(同质)的动物或样品;对于某病患者的受试对象而言,应当制定严格的纳入与排除标准;应尽可能采取双盲法或三盲法分配受试对象,尽可能减少或消除来自研究

者、特别是受试者心理的影响。对于环境因素的影响，应尽可能创造条件，使试验研究在特定的温度、湿度和光线条件下进行；对于仪器设备、试剂应尽可能校准，并经常对其进行检查和调试；对于指标的测定单位、测定方法应完全统一，实在不能统一的，应采取合适的方法进行换算；对于主要评价指标，若测定方法无法统一，应确定一个有资质的测定部门来完成测试，称为“中心试验室”，这是消除多个试验研究单位测定结果参差不齐的重要措施之一。

11.4 临床试验设计要点

临床试验是指在人体(患者或健康志愿者)中进行的试验药物、医疗器械或治疗方法的系统性研究，通过比较试验组与对照组的結果，揭示试验药物、医疗器械或治疗方法对人体的作用、不良反应，或探索药物在人体内的吸收、分布、代谢和排泄规律等的一种前瞻性研究。临床试验与试验室内进行的试验明显的区别是受试对象不同。试验室内进行的试验，受试对象通常是动物或标本，而临床试验的受试对象通常是某病患者。受试对象的变化，使临床试验设计变得更加复杂。本节将全面论述临床试验设计中不可忽视的重要问题。

11.4.1 临床试验中的法律与法规简介

新药临床试验在开始前，必须得到药品监督管理部门的批准，批准的标志是获得“新药临床研究批件”。临床试验的组织实施，也必须符合相关的法规要求。表 11-10 给出了中国、美国、欧盟和日本的管理当局、具体管理部门和主要相关法规。

表 11-10 中国、美国、欧盟和日本的管理当局、具体管理部门和主要相关法规

地 区	管 理 当 局	管 理 部 门	主 要 法 规
中国	国家食品药品监督管理局	国家食品药品监督管理局注册司	《中华人民共和国药品管理法》1984 年、药品临床试验管理规范(GCP)1999 年、《药品注册管理办法》2005 年
美国	Food and Drug Administration (FDA)	Center for Drug Evaluation and Research (CDER)	Food, Drug and Cosmetic Act, 1938; Code of Federal Regulations (CFR) title 21
欧盟	The European Agency for the Evaluation of Medicinal Products (EMA)	The Committee for Proprietary Medicinal Products (CPMP)	Directive 65/65/European Economic Community (EEC), 1965
日本	Ministry of Health and Welfare (MHW)	Central Pharmaceutical Affairs Council (CPAC) and the Pharmaceutical and Medical Device Evaluation Centre (PMDEC)	The Pharmaceutical Affairs Law, 1960

11.4.2 临床试验的特点与分期及伦理道德问题

1. 临床试验的特点

临床试验是以人为受试对象的研究，具有以下几个特点。

(1) 临床试验必须符合伦理道德规范，拟开展的研究项目必须经过有资质的伦理委员会的批准，研究者要与受试对象签订知情同意书。

(2) 临床试验需经历从探索到确证的递进性过程。

- (3) 临床试验的整个过程影响因素错综复杂, 其观测结果容易发生偏倚。
- (4) 对干预措施进行的前瞻性追踪研究。

2. 临床试验的分期

在进行药物临床试验之前应该进行新药的药理学和毒理学研究, 即临床前试验, 其内容包括药效学研究、一般药理学研究、药代动力学研究和毒理学研究等。这一阶段的工作主要在试验室和试验动物身上进行。临床试验分为 I、II、III、IV 期。新药在批准上市前, 通常应当进行 I、II、III 期临床试验; IV 期临床试验在新药上市后进行。

I 期临床试验: 初步的临床药理学及人体安全性评价试验。观察人体对于新药的耐受程度和药代动力学, 为制定给药方案提供依据。

II 期临床试验: 治疗作用初步评价阶段。其目的是初步评价药物对目标适应症患者的治疗作用和安全性, 也包括为 III 期临床试验研究设计和给药剂量方案的确定提供依据。此阶段的研究设计可以根据具体的研究目的, 采用多种形式。

III 期临床试验: 治疗作用确证阶段。其目的是进一步验证药物对目标适应症患者的治疗作用和安全性, 评价利益与风险关系, 最终为药物注册申请的审查提供充分的依据。试验一般应为具有足够样本量的随机、盲法、对照试验。

IV 期临床试验: 新药上市后由申请人进行的应用研究阶段。其目的是考察在广泛和长期使用条件下的药物的疗效和不良反应、特别是发生率很低但却属于严重不良事件的出现情况、评价在普通或者特殊人群中使用的利益与风险关系以及改进给药剂量等。

3. 临床试验的伦理道德问题

临床试验应遵守有关的法规和指南, 如药品管理法、药品注册管理办法、新药审批办法、药品临床试验管理规范(GCP)等。同时所有以人为受试对象的研究必须符合《赫尔辛基宣言》和国际医学科学组织委员会颁布的《人体生物医学研究国际道德指南》的道德原则。2008 年 10 月, 第 59 届世界医学大会通过《赫尔辛基宣言》修正案。修正案扩展了宣言的适用对象, 重申并进一步澄清了基本原则和内容, 加强了对受试者的权利保护, 同时还增加了临床试验数据注册和使用人体组织时的同意等新内容, 提高了人体医学研究的伦理标准。

进行药物临床试验必须有充分的科学依据。进行人体试验前, 必须周密考虑试验的目的、要解决的问题、预期的治疗效果及可能产生的危害。预期的收益应超过可能出现的损害。选择临床试验方法必须符合科学和伦理标准。

临床试验开始前, 试验方案须经临床研究负责单位的国家临床试验机构伦理委员会审议同意, 并签署批准意见后方能实施。在试验进行期间, 试验方案的任何修改均应经伦理委员会批准后方可执行。试验中发生任何严重不良事件, 均应向伦理委员会报告。

在药物和医疗器械临床试验的过程中, 必须对受试对象的个人权益给予充分的保障, 并确保试验的科学性和可靠性。研究者必须向受试对象提供口头或书面的有关临床试验的详细材料, 包括试验目的、预期收益、受试对象被分配到不同处理组而可能发生的风险与不便、因参加试验而受到损害或影响身体健康时能够获得的治疗和补偿。研究者不能强迫受试对象参加试验, 受试对象同意后, 须有受试者或其法定代理人在知情同意书上签字并注明日期, 执行知情同意过程的研究者或其代表也需在知情同意书上签名, 并注明日期。知情同意书应使用受试对象能够懂得的语言和文字, 受试对象保留在任何时候退出试验的权利。

11.4.3 如何制定出合理的诊断、纳入和排除标准

选择临床试验研究对象应该根据研究的目的选择。如果进行药物的疗效评价,应该选择患者作为研究对象,如果做预防措施的效果观察,就可以选择健康人。不管研究对象来自哪里,选择时都应该有统一的诊断标准、统一的纳入标准和统一的排除标准。

1. 如何制定出统一的诊断标准

好的诊断标准需要具备客观、量化的特点,最好选择国际公认的标准。但在大多数的情况下没有国际标准,这时应该选择已被国内同行认可的标准。如果国内外都没有统一的标准,就可以自己制定诊断标准,但必须尽可能地采用客观的标准,在操作时能够明确诊断。

在临床试验过程中,不管是选择病例也好,判断效果也好,诊断标准一经采用,就必须自始至终地去遵照执行,不可随意改动。

2. 如何制定出统一的纳入标准

诊断明确的病例不一定都符合研究的要求,应该制定纳入标准。纳入标准应该根据研究目的和实际情况制定,标准定得太高,增加工作量,不易找到研究对象。标准定得太低,又会影响研究结果。

在制定纳入标准时,应尽可能地选择对干预措施有反应的病例作为研究对象。一些旧病例、反复发作的病例,由于已经经过其他方法多次治疗,新的干预措施对它也不一定有效,因此尽量不要将其纳入。另外,在选择病例时,还需要考虑研究对象的代表性,选择的病例应体现这种疾病的特点,如果研究的某种疾病好发于老人,而选择的研究对象却是青年人,试验结果就难以说明问题。

3. 如何制定出统一的排除标准

除了制定纳入标准之外,还需要制定统一的排除标准。临床试验中,通常有下列情况的患者不能作为研究对象:①同时患另一种可能影响本试验效果的疾病的患者;②同时患其他严重疾病者,因为这样的患者可在研究过程中死亡或因病情恶化而被迫退出;③已知对药物有不良反应者。

4. 选择研究对象时应注意的其他事项有哪些

(1)入选的研究对象确能从科研中受益。如评价药物的疗效,研究人员应清楚地掌握该药的作用机制、适应证、禁忌证或敏感菌株等资料。这样就可选入敏感菌株感染的患者,而使患者受益,从临床试验角度来说也可获得阳性结果。

有关选择新旧患者问题应具体分析。一般来说在常见病、多发病的研究中要尽可能选择新发病患者作为研究对象,因为旧患者难以充分反映药物的疗效。若检验或估计新药的特殊疗效,可选择经多种方法久治未愈的患者,这样比较容易判断疗效,因为其本身就是一个历史对照。至于罕见病,因新病例数较少,在临床试验中不得不选入一些旧患者。

(2)研究对象的代表性。从临床试验出发,要求选入的研究对象在病型、病情以及年龄、性别等方面具备某病的特征,即代表性强。这样,试验获得的结论将具有明显的实用价值。若代表性差,科研结果的适用范围将受到限制。

若一项治疗措施的机制目前尚难肯定,或其机制不是单一的,可引起多种效应,此时应选择多质性的研究对象。如根据个人的具体情况进行体育锻炼以降低冠心病发病率或死亡率的临床试验,其机制可能是增强心脏功能,减轻体重,健全了意志或其他未知的作用所致。此时就要从不同性别、年龄、文化程度、职业、病种和病型的患者中选择研究对象,否则就不能全面反映体育锻炼的效应。

(3) 选择依从性好的患者作为研究对象。为了获得准确的结果, 研究人员通过观察和谈话以了解患者的情况, 从中选择那些能够服从试验安排并坚持合作的患者作为研究对象。若不依从患者的数量较大, 研究结果就会出现误差。

(4) 志愿者选作研究对象的问题值得商榷。在我国, 有些人出于崇高的目的和献身精神接受试验, 这是值得称赞的, 但这是少数。由于以下理由, 多位学者对这个问题持慎重态度, 即志愿者多有难言的苦衷, 如经济困难或自身患病求医心切等, 这些均可对研究结果的正确评价产生不利的影响, 甚至可出现假阳性应答; 再者, 志愿者代表性差。

11.4.4 如何确定主要疗效与安全性指标

(1) 主要疗效指标与次要疗效指标。主要指标又称为主要终点, 是与试验目的有本质联系的, 能确切反映药物有效性或安全性的观察指标。通常主要指标只有一个, 如果存在多个主要指标时, 应该在设计方案中, 考虑控制 I 类错误的方法。主要指标应根据试验目的选择易于量化、客观性强、重复性高, 并在相关研究领域已有公认标准的指标。主要指标必须在临床试验前确定, 并用于试验样本量的估计。次要指标是指与试验目的相关的辅助性指标。在试验方案中, 也需明确次要指标的定义, 并对这些指标在解释试验结果时的作用以及相对重要性加以说明。次要指标数目也应当是有限的, 并且能回答与试验目的相关的问题。如评价治疗 2 型糖尿病药物的有效性时, 糖化血红蛋白是评价其有效性的主要指标之一, 而空腹血糖和餐后 2 小时血糖是次要指标。对这些指标在解释试验结果时的作用及其相对重要性可在研究方案中加以说明。

(2) 安全性指标。临床试验中, 安全性评价是非常重要的一个方面。在安全性指标较多时, 也应指出主要与次要安全性指标分别是什么。在临床试验的早期, 这一评价主要是探索性的, 且只能发现常见的不良反应; 在后期, 一般可通过较大的样本进一步了解药物的安全性。后期的对照试验是一个重要的以无偏倚的方式探索任何新的潜在的药物不良反应的方法。药物安全性评价的常用统计指标为不良事件发生率和不良反应发生率。当试验时间较长、有较大的退出治疗比例或死亡比例时, 需用生存分析计算累计不良事件发生率。构成安全性评价的资料则主要来源于不良事件的临床表现、试验室检查等。从受试者中收集的安全性指标应尽可能全面, 主要包括受试者出现的所有不良事件的类型、发生时间、严重程度、处理措施、持续的时间、转归等。所有的安全性指标在评价中都需十分重视, 其主要分析方法需在研究方案中指明。

11.4.5 如何遵循随机、对照、重复和均衡原则

1. 如何遵循随机原则

随机化是使临床试验中的受试者有同等的机会被分配到试验组或对照组中, 而不受研究者和/或受试者主观意愿的影响, 可以使各处理组的各种影响因素(包括已知和未知的因素)分布趋于相似。随机化包括完全随机、分层随机、动态随机, 与盲法合用, 有助于避免因处理分配的可预测在受试者的选择和分组时可能导致的偏倚。

临床试验中可采用分层、区组随机化方法。分层随机化有助于保持层内的均衡性, 特别在多中心临床试验中, 中心就是一个分层因素。另外为了使各层趋于均衡, 避免产生混杂偏倚, 按照基线资料中的重要预后因素(如病症的严重程度)等进行分层, 对促使层内的均衡安排是很有价值的。区组随机化有助于减少季节、疾病流行等因素对疗效的影响。区组的大小要适当, 太大容易造成组间不均衡, 太小则容易造成同一区组内受试者分组的可猜测性。研究者及其有关人员应对区组的大小保持盲态。

当样本量、分层因素及区组大小决定后, 由参与临床试验的统计学专业人员在计算机上使用统

计软件产生随机分配表。临床试验的随机分配表就是用文件形式写出对受试者的处理安排,即处理(或在交叉试验中的处理顺序)的序列表。随机分配表必须有可以重新产生的能力,即当产生随机数的初值、分层、区组决定后能使这组随机数重新产生出来。

完全随机化的利弊:完全随机化实施起来很方便,在样本含量超过一千以上时,其随机化效果会比较理想;而当样本含量在几百内、尤其在几十之内时,有时,完全随机化的效果很糟糕。糟糕的标志是:试验组与对照组之间在某些重要非试验因素方面很不均衡。解决问题的方案是:找准找全对观测结果可能有影响的重要非试验因素,使它们形成复合型区组因素,据此对全部受试对象进行分层随机化。

分层随机化的实施:若考察某试验药物的降血脂作用,进行Ⅱ期临床试验,通过样本含量计算得到试验组和对照组各需做120例,共240例。受试对象应属于高血脂患者,此时,随机分组的结果应确保试验组与对照组患者的基础平均血脂基本相同,若分组结果离此要求相差甚远,就应按血脂测定结果对受试对象进行分层随机设计。由于临床试验通常由多个试验中心来共同完成,“试验中心”就成为一个不可回避的重要非试验因素。如果年龄、性别对观测结果的影响也不可忽视,应将在同一个试验中心接受试验,且性别相同、年龄和血脂含量分别相近的受试对象视为“一层”,这样,分层随机便可使试验中心、性别、年龄和血脂含量对试验组与对照组的影响达到均衡一致。此时,可以称“分层因素”为由四个重要非试验因素复合而成的“复合型区组因素”。将每个区组内的“同质”受试者完全随机地均分入试验组与对照组中。产生240例受试者所接受处理的随机数字,即列出流水号为001~240例所对应的治疗分配方案(即随机编码表)。若整个过程由6所医院协同完成,则每个临床试验中心有40例,试验组和对照组各20例。

若整个临床试验需要在一个较长时间内才能完成,则“季节”和“疾病流行周期”也许有必要作为“重要非试验因素”,应当一并纳入“复合型区组因素”,以减少季节、疾病流行等因素对疗效的影响。此时,每个试验中心完成的40例受试者应均匀地分布在全部试验阶段内。

动态随机化的实施:根据专业知识选取几个必须加以控制的重要非试验因素,假定一个是患者的“性别(分为男、女)”,另一个是患者的“病情(分为轻、中、重)”。将先来的两位患者在试验组与对照组各放一人,记下他们的性别和病情,记分的方法是每个因素的每个水平出现一次记1分,计算两组各因素对应水平的得分之差的绝对值,最后求出绝对值之总和,称此“和”为两组患者在两个重要非试验因素上的不平衡指数。若再来第三位患者,分别依次将此患者放入试验组、对照组各一次,每次都根据他(或她)的性别、病情累加到原有患者的基础之上,可以得到两个不平衡指数,取不平衡指数最小的那种分组方法,这样第三位患者的分组就定下来了,用同样的方法去分配以后的该病患者,直到两组有了事先规定的样本含量时就停止。现举例说明如下。

假定在对肩周炎患者分两组时考虑病情(轻、中、重)、患病时间(短、长)和日常运动量(少、多)三个重要的非试验因素,再假定治疗组和对照组各有了一位患者,他们的基本情况如表11-11所示,第3位来的是一位新患者,病情重、患病时间短、日常运动量多,问将此新患者分入哪个组中去为好?

试着将新患者分别分入治疗组和对照组,取“不平衡指数”较小者所对应的分配方案,分别参见表11-12和表11-13。

因5<7,故本例应取表11-12作为分配结果,若再有新患者,分配方法相同。

2. 如何遵循对照原则

为了评价一个新药的疗效和安全性,必

表 11-11 两例患者的基本情况及分组后的平衡程度

试验 分组	例 数			
	病情 (轻 中 重)	患病时间 (短 长)	运动量 (少 多)	合计
治疗组	0 1 0	0 1	1 0	
对照组	1 0 0	1 0	1 0	
差的绝对值	1 1 0	1 1	0 0	4

注:表中合计值“4”就是“不平衡指数”,下同。

须有供比较的对照组。对照组与试验组唯一的差别是试验组接受新药治疗, 对照组则接受对照药物的治疗。设立对照组的主要目的是判断受试者治疗前后的变化(如体征、症状、死亡、复发、疗效、不良反应等)是由试验药物, 而不是其他因素(如病情的自然发展过程或者受试者机体内部环境的变化)引起的, 对照组的设置还能科学地回答未服用试验药物会发生什么情况。

表 11-12 将此名新患者分入治疗组后的平衡程度

试验 分组	例 数			
	病情 (轻 中 重)	患病时间 (短 长)	运动量 (少 多)	合计
治疗组	0 1 1	1 1	1 1	
对照组	1 0 0	1 0	1 0	
差的绝对值	1 1 1	0 1	0 1	5

表 11-13 将此名新患者分入对照组后的平衡程度

试验 分组	例 数			
	病情 (轻 中 重)	患病时间 (短 长)	运动量 (少 多)	合计
治疗组	0 1 0	0 1	1 0	
对照组	1 0 1	2 0	1 1	
差的绝对值	1 1 1	2 1	0 1	7

临床试验要求试验组和对照组来自相同的受试者总体。不但在试验开始时两组受试者基本情况是相同的或相似的, 而且在试验进行中除了试验药物不相同外, 其他条件均保持均衡。因此在实际选择对照时, 应尽量选择公认有效的对照药, 并且是进入国家标准的药物。

另外, 在实际的新药临床试验中, 试验药可设置一个或多个对照组, 试验药也可设多个剂量组。对照组可分为阳性或阴性对照。阳性对照一般采用按所选适应症的当前公认的有效药物, 阴性对照一般采用安慰剂, 但必须符合伦理学要求。试验药物设一个或多个剂量组完全取决于研究者对试验药的疗效和安全性的把握程度。若通过临床前期的动物试验或Ⅰ期临床试验, 已经比较准确地把握了此试验药物的合适剂量, 则进行正式的Ⅱ期和Ⅲ期临床试验时, 试验药物可仅取一个特定的剂量; 否则, 需通过多设置几个试验药物的剂量组进行量效关系的探索性研究(称为ⅡA期临床试验), 然后, 再选取一个比较合适的剂量组进行正式的Ⅱ期临床试验(称为ⅡB期临床试验), 而Ⅲ期临床试验一般不再进行量效关系的探索性研究了。

同样, 在医疗器械临床试验中, 也需要设置对照组。在医疗器械临床试验中, 可为试验器械设置一个或多个对照组, 也可按照试验器械的若干治疗强度/方法设立对照组。对照器械的选择应符合“对等性”要求(例如, 新式止血纱布与已通过国家级评审机构评价的老式止血纱布比较, 若与老的止血粉比较, 就不对等了)。对照组可分为阳性或阴性对照。阳性对照一般采用适用于所选适应症的、当前临床广泛使用的、公认的有效器械; 阴性对照一般采用没有经临床证实疗效的安慰器械, 但也必须符合伦理学要求。另外, 医疗器械临床试验有时也可采用历史对照, 但采用历史对照时, 历史对照组由一组与试验器械组患者具有相同疾病和严重程度、非同时进行同类器械治疗的患者组成。但是, 值得注意的是: 历史对照难以保证研究人群具有可比性, 如不同时期相同疾病的人群在治疗水平、经济状况、人种分布与构成、年龄分布与性别构成、平均发病时间、疾病严重程度构成等方面不同, 即使得到阳性的结果也很难保证是研究器械的有效所致, 因此, 历史对照应慎用。

3. 如何遵循重复原则

重复原则是指研究的试验组与对照组中应有一定数量的受试对象, 即试验单位数要达到一定的数量。在试验设计中遵循重复原则的一个重要标志是如何科学合理地确定样本含量。一般来说, 抽样误差随着独立重复试验次数的增多而减小, 观测结果的可信度也逐渐增高。混杂因素的影响也可以通过一定数量的重复而部分抵消, 从而增强组间的可比性。但是试验条件会因为样本含量过大导致控制难度增加, 可能引入非随机误差, 给观测结果带来偏性。

在临床试验中, 样本含量确定的原则是: 如果经过统计学计算, 样本含量少于《药品注册管理办法》的规定, 按照国家要求确定; 否则, 按照统计学计算结果确定。

我国《药品注册管理办法》对临床试验的最低病例数(试验组)的要求如下。

(1) 临床试验的病例数应当符合统计学要求和最低病例数要求。

(2) 临床试验的最低病例数(试验组)要求: I 期为 20 ~ 30 例, II 期为 100 例, III 期为 300 例, IV 期为 2000 例。

(3) 属注册分类 1、2、4、5、6 的新药, 以及 7 类和工艺路线、溶媒等有明显改变的改剂型品种, 应当进行 IV 期临床试验。

(4) 生物利用度试验一般为 18 ~ 24 例。

(5) 避孕药 I 期临床试验应当按照本办法的规定进行, II 期临床试验应当完成至少 100 对 6 个月经周期的随机对照试验, III 期临床试验应当完成至少 1000 例 12 个月经周期的开放试验, IV 期临床试验应当充分考虑该类药品的可变因素, 完成足够样本量的研究工作。

(6) 新的中药材代用品的功能替代, 应当从国家药品标准中选取能够充分反映被代用药材功效特征的中药制剂作为对照药进行比较研究, 每个功能或主治病证需经过 2 种以上中药制剂进行验证, 每种制剂临床验证的病例数不少于 100 对。

(7) 改剂型品种应根据工艺变化的情况和药品的特点, 免除或进行不少于 100 对的临床试验。

(8) 仿制药视情况需要, 进行不少于 100 对的临床试验。

(9) 进口中药、天然药物制剂按注册分类中的相应要求提供申报资料, 并提供在国内进行的人体药代动力学研究资料和临床试验资料, 病例数不少于 100 对; 有多个主治病证或适应症的, 每个主要适应症的病例数不少于 60 对。

不同的试验设计类型(如配对设计、成组设计或两因素析因设计)和比较类型(如非劣效性检验、差异性检验、等效性检验和优效性检验)对应的样本含量计算公式不同。但总体来说, 决定样本含量大小主要取决于以下几个因素: ①可信度($1 - \alpha$); ②把握度($1 - \beta$); ③总体均值 \bar{x} 和标准差 σ 、总体率 π 或总体相关系数 ρ ; ④容许误差 Δ , 即当样本含量为由合适的公式计算所得的 n 值时, 用关键的样本统计量(如 \bar{x})来估计总体参数(如 μ)时, 二者之差不超过 Δ 的概率是 $1 - \alpha$; ⑤三种特殊比较类型下的检验界值(δ)。

另外, 还应根据临床试验药物本身的特点考虑脱落率所带来的影响。

4. 如何遵循均衡原则

均衡是指试验组与对照组之间除了所用的药物不同外, 其他任何可能影响观测结果的因素在组与组间均衡一致。

在试验设计时, 有时因样本含量不够大, 完全随机分组的结果并不能确保两组受试对象在某些重要非试验因素方面均衡一致。因此, 有必要强调均衡原则, 即在进行试验分组时, 必须设法确保试验组与对照组之间在所有重要非试验因素方面均衡一致。例如, 若考察“A 药”的降血脂作用, 受试对象应属于高血脂患者, 此时, 随机分组的结果应确保试验组与对照组患者的平均血脂基本相同, 若分组结果离此要求相差甚远, 就应按血脂测定结果对受试对象进行分层随机设计。如果年龄、性别对观测结果的影响也不可忽视, 应将性别相同、年龄和血脂含量分别相近的受试对象视为“一层”, 这样形成多个层, 再将每层中的受试对象随机地均分入试验组与对照组, 分层随机便可使性别、年龄和血脂含量对试验组与对照组的影响达到均衡一致。

值得注意的是, 在新药和医疗器械临床试验中, “疾病类型”、“病情轻重”和“患病时间长短”常常是不可忽视的重要非试验因素。与此同时, 还应记录下每位受试者的定量影响因素的具体取值(如具体年龄、血脂含量、雌激素含量, 如有可能, 还应记录与血脂有关的其他定量影响因素的具体取值, 如胆固醇含量等), 在进行统计分析时, 可采用相应设计类型下“协方差”分析方法校正已观测的定量影响因素的干扰和影响, 更真实地显露出处理因素取不同水平对观测结果影响之间的差异。

11.4.6 如何实施盲法与降低偏性及多中心临床试验的中心效应

1. 如何实施盲法

盲法(Blind Method)作为纠正偏倚的一个重要措施,特别在临床试验中,是为了避免研究者与受试者的主观和心理因素对试验结果的干扰。临床试验根据设盲的程度分为单盲、双盲、三盲和非盲(开放性试验)。仅受试者不知所接受的是何种疗法时,称单盲(Single Blind);而参与临床试验和判断疗效的医护人员与受试者都不知任何一位受试者所接受的为何种疗法时称双盲(Double Blind),从而使研究结果得以客观评价;三盲(Triple Blind)则是研究者(包括医护人员)、研究对象和负责资料收集和分析的人员均不知任何一位受试者所接受的为何种疗法,从而更好地避免了偏倚。

除开放性试验(Open Label)和某些不宜设盲的试验如外科手术、引起生活方式改变等干预试验外,一般均应采用盲法。采用单模拟或双模拟等措施,能够保证试验过程中盲法的实施,并且是最有效的施盲措施。双模拟技术(Double Dummy)指由申办者制备一个与试验药外观相同的安慰剂,称为试验药的安慰剂;再制备一个与对照药外观相同的安慰剂,称为对照药品的安慰剂。也就是说,试验组服用试验药加上对照药的安慰剂;对照组服用对照药加上试验药的安慰剂,各药及其安慰剂服用方法相同。从整个用药情况来看,每个病例所服用的药物,每日次数,每次片数都是一样的,这就保证了盲法的实施。例如,某研究,试验药为黄连素,对照药为阿莫西林,则黄连素组应配与阿莫西林外观一样的模拟剂,而阿莫西林组应配与黄连素外观一样的模拟剂,像这种两组互相模拟的试验,即为双模拟,整个盲法的设置称为双盲双模拟。

2. 如何降低偏性

临床试验中的偏倚一般包括选择偏倚、信息偏倚和混杂偏倚。

选择偏倚是在临床试验中,由于选择的试验对象或观察指标不恰当而引起的偏倚。这种偏倚使得从样本得到的结果在推广到总体时出现系统的偏差。

降低选择偏倚的有效方法是,在临床试验设计阶段对产生选择偏倚的原因采取相应的措施,防止偏倚产生。在实施阶段,要严格按照研究方案规定的条件纳入受试者,对临床试验对象要随机分组,必要时采用分层随机分组。尽量要使两组除研究的试验因素取值不同外,其他的有关条件和影响因素均衡一致。

信息偏倚是由于在临床信息收集、整理过程中各种原因的影响而出现的误差。它可能来自临床试验的全过程。

控制观察性偏倚的产生,主要是在收集信息阶段,针对产生信息偏倚的原因采取相应措施。加强对临床试验人员的试验前培训,提高临床试验质量;采取盲法收集资料和数据,使临床研究人员、受试者在不知道分组的情况下对试验效应作出客观、公正的测量或评价,对临床试验所用的测量仪器也要在使用前加以校正,应检查试剂是否过期,并严格执行标准操作规程。

混杂偏倚是指当研究某一处理因素与疾病的疗效关系时,另一种伴随的非试验因素产生的效应,干扰着试验因素所产生的效应,这一伴随因素称为混杂因素。

降低混杂偏倚,在试验设计阶段可以按照可预期的混杂因素进行分层随机化;在统计分析阶段,若混杂因素是定性的,可将其与试验因素进行联合考虑,形成多因素试验设计类型,并选用相应设计下的统计分析方法处理;若混杂因素是定量的,则可将其当作协变量,采用相应设计定量资料的协方差分析。

3. 如何降低多中心临床试验的中心效应

多中心试验是指由一个单位的主要研究者总负责,多个单位的研究者合作,按同一个试验方案

同时进行的临床试验。多中心试验可以在较短的时间内入选所需的病例数,且入选的病例范围广,临床试验的结果更具代表性。但影响因素也随之更复杂。

多中心临床试验有时会出现中心效应(即不同中心的观测结果之间差异有统计学意义),因此,进行多中心试验必须在统一的组织领导下,遵循一个共同制订的试验方案完成整个试验。各中心试验组和对照组病例数的比例应与总样本的比例相同,以保证各中心齐同可比。多中心试验要求各中心的研究人员采用相同的试验方法,试验前对人员统一培训,试验过程要有监控措施,以保证他们的诊断、观测、评价上的一致性。当主要指标可能受主观影响时,需进行统一培训和一致性检验,在正式开展试验之前,应设法使之达到符合一致性的要求。当主要指标在各中心的试验室的检验结果有较大差异、参考值范围不同或度量单位不同时,应采取相应的措施,如统一由中心试验室检验、采取相同的参考值范围、相同的度量单位或推导出校正公式。在双盲多中心临床试验中,盲底是一次产生的,应按中心分层随机;当中心数较多且每个中心的病例数较少时,可统一进行随机,不按中心分层。

11.4.7 如何正确把握临床试验中的四种比较类型

临床试验中的比较类型,按照研究目的可分为差异性试验、优效性试验、等效性试验和非劣效性试验。

差异性试验是指主要研究目的为显示两种治疗效果之间的差别大小与0之间的差别是否具有统计学意义的试验。在试验设计阶段不需要设定任何界值。

等效性试验是指主要研究目的为显示两种治疗效果之间的差别大小在临床上并无重要意义的试验。在试验设计阶段需要设定等效性界值(δ_L, δ_U)来界定两种治疗的等效性。

优效性试验是指主要研究目的为显示试验药的治疗效果优于对照药(安慰剂对照或阳性对照)的试验。在试验设计阶段需要设定一个界值 δ_U 来界定试验药的优效性。

非劣效性试验是指主要研究目的为显示试验药的治疗效果在临床上不比阳性对照药差的试验。在试验设计阶段需要设定一个界值 δ_L 来界定试验药是否不比阳性对照药疗效差过预先设定的这个界值。

11.5 调查设计要点

调查性研究又称为观察性研究或非试验性研究,它对受试对象不施加任何干预措施,是在完全“自然状态”下对研究对象的特征进行观察、记录,并对观察结果进行描述和对比分析。调查性研究广泛应用于医学、社会学、教育学等领域,人口普查、市场调查、民意测验等都是调查性研究的具体应用。一般来说,涉及人的行为目的、观点、态度、信念、期望、疾病状况等方面的研究适合采取调查性研究的方法,在医学中,调查性研究还可用于病因的探讨。

11.5.1 如何正确把握调查设计中的几个关键环节

调查设计是实现调查研究目的的重要保证,它是科学合理地安排调查的资料收集、整理和分析的各个环节的集中体现。其实施步骤一般分为:①明确调查研究目的和调查指标;②确定调查对象并估计样本含量;③明确调查方法和调查方式;④拟定调查项目和调查表;⑤制订调查组织计划;⑥资料的整理和分析。

1. 明确调查研究目的和调查指标

明确调查研究目的是整个调查研究中最核心的问题。调查目的就是明确要解决哪些问题。医

学领域中的研究目的主要是描述人群疾病和危险因素状况、探讨疾病病因学和流行病学的原理、研究疾病的防控效果等。调查目的应具有具体、实用、可行的特点,一次研究应有一个明确的研究目的,同时也可以兼顾几个次要的研究目的。例如,某食管癌高发区危险因素调查,其主要研究目的是调查该地区食管癌发病的相关影响因素,次要目的是为疾病防治工作提供依据,并为病因研究提供线索。

调查者应根据调查目的制定调查指标,调查目的是选定调查指标的依据,调查指标又是调查目的的具体体现。调查指标要精选,重点要突出,尽量选用客观性强、灵敏度高、精确度好的定量指标。如上例关于食管癌调查的调查指标可以确定如下:①该地区最近几年不同性别、不同年龄段的食管癌发病率;②食管癌相关危险因素的暴露水平,如微量元素和维生素 C 水平测量、霉变食物、酸性食物等摄入及饮酒情况等。

2. 确定调查对象并估计样本含量

要根据调查目的和调查指标来确定调查对象,即划清调查总体的同质范围。明确的纳入标准和排除标准可以用来明确调查对象,同时也是控制选择偏倚的有效手段。调查总体中的观测单位可以是一个人、一个家庭或一个群体。以人为观察单位的调查较多,例如,河南某村庄未婚成年男子的艾滋病感染情况调查,调查对象是该村常住居民中的未婚成年男子,纳入和排除标准应该明确调查对象的性别、年龄、户籍、婚姻状况等标准。

为了保证调查研究的质量,还必须保证足够的观测单位数量,这就是样本含量的问题。样本含量的大小与抽样误差有直接关系,因此确定一个恰当的样本含量可以将抽样误差控制在可接受的范围。一般来说,样本含量过大,会增加实际工作的难度,而且很可能会增加非抽样误差,降低调查研究质量;样本含量过小,所得指标变异较大,总体估计精度差,检验效能较低,很难获得正确的研究结果。因此,样本含量估计是在保证一定的参数估计精度和检验效能的前提下,所需要的最少样本数。调查研究的样本量估计需要的信息可以通过查阅文献、实施预调查等加以掌握,而且根据分析目的、资料类型、抽样方法等有所不同。

3. 明确调查方法和调查方式

研究者可以根据不同的调查目的和实施条件选择合适的调查方法。根据调查范围大小,调查方法可分为普查、抽样调查、典型调查;根据研究时间,又可以分为横断面研究和纵向研究;按抽取样本的方式,可分为概率抽样调查和非概率抽样调查。

普查是指对调查范围内的全部观察对象进行调查,这种调查成本较高,实施难度大,而且系统误差和过失误差往往比较大。抽样调查是医学科研中最常用的调查方法,它只对总体中有代表性的部分观察对象进行调查,因此可以节约调查成本,并可以获得较为细致准确的信息。典型调查也称为案例调查,即从调查总体中选择典型的观察对象进行调查,这种方法有利于对事物现象的规律做深入研究。横断面研究是指在特定的时间点或相当短的时间内,对观察对象进行调查研究。纵向研究是指对事物现象发展变化的动态、持续的观察过程,根据研究方向可以分为回顾性研究(如病例对照研究)和前瞻性研究(队列研究),前者是由果推因,后者是由因溯果。

无论采用哪种方法进行调查,获取资料的方式主要有直接观察法和采访法两种。直接观察法是由调查员直接对观察对象进行检查、测量、提问、记录取得资料。该方法取得的信息较为真实可靠,但是调查所需成本较高;采访法分为直接和间接法,它们包括多种组织方式,如面对面、自填问卷、电话采访、电子邮件采访等。各种采访方式都有各自的优缺点,如面对面采访交流沟通方便,准确性和应答率较高;自填问卷节约调查成本,保密性较好,但是应答率较低;电话采访和电子邮件采访应答率更低,调查质量更差,适用于调查内容简单、调查质量要求较低的调查。

4. 拟定调查项目和调查表

根据调查指标确定调查项目，即调查的具体内容，包括分析项目和备查项目。分析项目是直接用于计算调查指标以及考虑因素之间关系所必须的内容；备查项目是为了便于核查、填补和更正而设置的，通常不直接用于分析，如姓名、联系方式等。

把调查项目按逻辑顺序列成表格形式供调查使用就是调查表。调查表应精简，其中分析项目一个也不能少，备查项目不宜多。每个调查表只调查一个对象时用单一表或卡片，项目可以较多；每个调查表调查多个对象时可用一览表，适用于项目较少的情形。

5. 制订调查组织计划

调查研究是一项社会性很强的工作，严密的项目实施组织计划有利于保证研究工作顺利开展。调查的组织计划应包括组织领导、宣传动员、时间进度、分工协调、经费预算、运作管理、现场组织、调查员培训、调查表核查制度、资料汇总整理要求等。大规模的调查研究往往需要进行小范围的预调查，以便修改完善组织计划和调查设计方案。

6. 资料的整理和分析

资料的整理和分析包括审核复查、录入整理、统计分析这三个过程。①资料审核复查：审核是指要对调查收集的资料进行审查核对，主要是包括问卷的完整性、逻辑性检查。复查是指从所调查的样本中随机抽取一定比例的被调查者进行回访，以检验调查的质量。②资料录入整理：调查的资料录入可以通过 EpiData 等录入软件来进行，可以采取双人双录的方式，即两名数据录入人员分别独立进行输入，校对数据库并消除录入误差。数据整理：这个过程一般采用计算机来实现，主要包括对数据的清理、分组和汇总，以便进一步的统计分析。③统计分析：应根据分析目的、抽样方法、资料类型以及变量属性等因素，选用合适的统计分析方法。

11.5.2 常用抽样方法

调查研究中研究者常常采用随机抽样的方法获得样本，这些方法具有统计学理论依据，因此可以对抽样误差进行计算，从而对调查结果的准确性进行评估。常用的随机抽样方法有单纯随机抽样、分层抽样、整群抽样、系统抽样以及多阶段抽样。

进行随机抽样时，首先需要确定抽样总体和抽样框。抽样总体是根据研究目的确定的具有同质性的抽样单位的全体，但总体往往界定起来比较模糊，因此在实际应用中通常通过抽样框来界定。抽样框是包含全部抽样单元的目录性清单，在抽样框中，每个抽样单元都有自己独一无二的位置，这常常通过编号来实现。抽样框的确定在抽样调查过程中极为重要，然而确定抽样框有时也比较困难，它必须包含抽样总体中的所有抽样单位，否则抽取的样本就存在严重的偏倚，调查结论也将不可信。例如，调查者采用电话采访的方式对某大学的全体学生进行抽样调查，把所有学生的手机号码组成的清单作为抽样框，然后通过计算机随机选取号码作为样本。由于并非所有学生都在使用手机，没有手机的学生并未包含在抽样框中，也就是说调查总体中不是所有学生都能被选中，因此这就存在选择偏倚。

下面介绍几种常用的随机抽样方法。

(1) 单纯随机抽样：又称为简单随机抽样，它是利用随机的方法从总体中抽取部分抽样单位作为样本，总体中每个抽样单位被抽中的概率相等。单纯随机抽样可以采用掷硬币、掷骰子、抽签或者查随机数字表的方法，也可以在计算机上利用 SAS、SPSS 等软件来实现。单纯随机抽样是最简单的一种随机抽样方法，也是其他抽样方法的基础。

单纯随机抽样的优点是简便易行，缺点是不适合样本量很大的研究，样本量很大时对每个观察

对象编号工作量大,有时很难做到;这种方法也不适用于那些个体差异很大的研究对象的抽样,因为在这种情况下,需要在样本量很大时才具有代表性。

(2)系统抽样:也称为等距抽样或机械抽样,它是按照一定的顺序对总体中各抽样单位进行排序,根据样本大小确定抽样间隔,然后随机确定起点,按照抽样间隔抽取每一个单位。它也是在随机抽样的基础上进行的,即要求每次抽样的起点必须是随机的。该方法操作方便,若对研究总体排序标志有所了解时,可利用某些规律达到较好的抽样效果,故抽样误差通常小于单纯随机抽样。缺点是当总体的观察单位按顺序有周期趋势或单调增(减)趋势时,容易产生系统误差,所以该法适用于观察单位分布较均匀的总体。

(3)分层抽样:是先按对观测指标影响较大的某种特征,将总体分为若干个层,再从每一层内随机抽取一定数量的抽样单位组成样本。分层原则是使层内差异尽可能小,层间的差异尽可能大,从而提高调查的精度。分层抽样按照各层之间的抽样比是否相同,可分为等比例抽样与非等比例抽样。分层抽样的特点是:①调查的精度比较高,因为合理的分层提高了层内同质性,减少了层内的抽样误差;②可以对不同的层采用不同的抽样方法,抽样方法比较灵活;③还可以对各层进行独立分析。

(4)整群抽样:又称为集团抽样,它是先将总体分成 K 个群(组),从中随机抽取 k 个群,然后把抽取的群中的所有观察单位组成样本。在实际调查中,经常用到自然群,如省市、村镇、企业、学校等,要求群间的变异越小越好。整群抽样的优点是便于组织实施,由于调查的观测单位比较集中,可以大大减少调查成本,容易控制调查质量。缺点是抽样误差一般大于单纯随机抽样,为了降低抽样误差,应该尽量缩小群间的差异,增加总体的群数,同时增加被抽取的群数以扩大样本量。例如,某个调查总体是某中学初一全年级共 800 名学生,每班大约 40 人,调查者想了解学生平均每周零食花费情况。若以班级为单位进行整群抽样,可以从初一所有班级中随机选取 5 个班,然后对 5 个班中约 200 名学生进行调查。

(5)多阶段抽样:又称为多级抽样,它是指抽样过程分为多个阶段来进行,每个阶段可以采取不同的抽样方法。多阶段抽样先从总体中抽取一级抽样单元,再从中抽取范围较小的二级抽样单元,然后依次抽取更次级的抽样单元,直到获得所需要的样本量时抽样才结束。多阶段抽样的优点是组织实施起来比较方便,尤其是不需要对每一级抽样单位编制完全的抽样框。缺点是抽样误差的估算比较复杂。例如,某调查项目想了解某省所有区县级医院的病床使用情况,先将该省所有医院按照其地域分布关系分为东西南北中五个区域,每一个区域中随机抽取一个城市(一级抽样单位),然后从已经抽取的城市中随机抽取若干个区县级医院(二级抽样单位),被抽到的区县的全部医院接受调查,这就属于多阶段分层整群抽样过程。

(6)非概率抽样:以上 5 种抽样方法都考虑了等概率的原则,即抽样框中的成员或单位被抽中的概率相同。如果对总体无法确知,只能采用非概率抽样方法,它主要依据研究者的意愿、判断或方便程度等条件来抽取调查对象,因不考虑随机抽样中的等概率原则,往往产生较大的抽样误差,难以保证样本的代表性。常用的非概率抽样方法有立意抽样、偶遇抽样、雪球抽样等。

①立意抽样,又称为目的抽样或判断抽样,是研究者根据研究目标和对情况的主观判断来选择和确定调查对象的方法。假如欲调查吸毒者吸毒的原因,由于吸毒者较少,吸毒行为又极其隐蔽,无法判断其总体有多大,无法采用随机抽样的方法,只能在吸毒人群较多的地区进行调查,或选择对吸毒者有档案管理的社区进行。

②偶遇抽样,又称为便利抽样或方便抽样,是指调查者根据实际情况使用对自己最为便利的方式来选取样本,可以是抽取偶然遇到的人或选择那些容易找到的人作为调查对象。例如,某医生要调查患者对目前医疗收费的看法,直接选择到医院看病的患者进行调查就是便利抽样。

③雪球抽样,当我们无法了解总体情况时,可以从能找到的少数个体入手,对他们进行调查,

并请他们介绍其他符合条件的人,扩大调查面,如此重复下去,直到达到所需的样本含量。例如,吸毒者、女性性服务工作者的调查可以采用雪球抽样的方法。

11.5.3 如何进行调查质量控制

调查研究中调查结果常常出现误差,误差大致分为两类,即抽样误差和非抽样误差(包括系统误差和过失误差)。抽样误差在抽样调查中是不可避免的,它主要受到总体内个体之间的离散程度、抽样方法以及样本大小等因素的影响。抽样误差有一定的规律,因此可以运用统计学方法进行估计。引起非抽样误差的原因较多,也比较复杂,控制起来较困难。问卷设计不合理、仪器测量不精确、调查人员操作不当、询问方式不正确、资料的录入和汇总计算有误等,都可以造成明显的非抽样误差。因此,对非抽样误差的控制应贯穿研究的整个阶段。本小节重点讨论常见的几种非抽样误差及其控制方法。

1. 如何进行研究各阶段的质量控制

(1)设计阶段:该阶段特别要注意明确研究范围和对象、选择恰当的调查指标、设计完善的调查表、采取合理的调查方式、开展必要的预调查等。

(2)调查阶段:该阶段要求严格按照研究设计方案来执行,要重视调查人员的培训,提高其调查技巧和调研水平,最大程度地争取调查对象的配合,提高问卷的应答率,降低误答率。

(3)资料整理和分析阶段:应采取有效的措施减少数据录入、汇总和计算等方面的错误,加强问卷审核、数据核查、结果复核等质量控制工作。

2. 如何进行非抽样误差的控制

非抽样误差是指各种抽样误差以外的因素造成的测量值与真实值之间的误差,也称为偏倚。常见的偏倚有选择性偏倚、信息偏倚和混杂性偏倚。

(1)选择性偏倚:即由于在选择调查对象的方式上不正确造成的系统误差,调查研究中的选择偏倚主要是由于抽样框误差造成的。如果抽样框设计不合理,造成抽样框与目标总体不一致,这时无法保证样本的代表性,从而使调查结果存在偏差。产生抽样框误差的原因主要有:①抽样框未包含所有目标总体的观测单位;②抽样框包含了非目标总体的观测单位;③复合连接,这种情况是抽样框观测单位与目标总体观测单位不符合一对一连接的原则。例如,对某市曾入院治疗的肺心病患者进行抽样调查,若以该市肺心病患者的住院病历作为抽样框,因为很可能有的患者曾在多家医院入院治疗,这些患者在抽样框中被抽中的概率比只入住过一家医院的患者大,因此,获得的抽样样本存在选择性偏倚。

(2)无应答偏倚:是指由于各种原因导致不能获得某些调查对象的资料信息,造成这种情况的原因有:①与调查对象无法取得联系;②调查对象拒绝参与调查;③问卷中部分问题出现拒答、忘答等;④其他不能完成调查的客观因素,如文盲、弱智、疾病等。

(3)测量偏倚:是指调查研究过程中观察和测量造成的偏移,它对调查质量影响比较大。造成测量偏倚的原因有很多,调查研究中调查对象往往容易提供不正确的信息,例如对问卷中问题的理解误差、记忆误差、刻意隐瞒或其他影响问卷质量的因素。此外,调查设计方案是否合理、调查员操作是否恰当都能造成测量误差。

(4)混杂偏倚:在对比组之间,除了研究者拟考察的因素分别取不同水平外,还受一个或多个其他因素不平衡的影响,后者就称为混杂因素,由混杂因素所导致的错误信息就称为混杂偏倚。例如,在研究吸烟与肺癌的联系中,性别和年龄都是混杂因素,因为性别和年龄与吸烟均有联系,它们在吸烟者与不吸烟者中的分布不相同。若在调查设计和统计分析阶段不采取有效措施,则性别和年龄会歪曲吸烟对肺癌的作用。

11.5.4 如何编制一份具有可操作性的调查表

1. 调查表的结构

调查表一般包括：①问卷说明，包括调查目的、保密声明、注意事项等内容；②核查项目，包括问卷编号、调查日期、复核结果、调查员和审核员签名等，这部分是用于质量控制，保证调查表填写的完整和准确；③调查项目，包括一般情况调查（调查对象的背景资料）、问卷主题内容、问卷项目编码等。

2. 调查表制定步骤

①拟定调查表的内容纲要：根据调查目的和调查计划，拟定调查表的调查纲要，明确调查主题和调查项目。②编写问卷及其顺序安排：主要是确定问卷结构、编制调查问题、问题的筛选和编排。③预调查及修改：将拟定好的调查表在小范围内进行预调查，并初步评价调查表的合理性，并做必要修改完善。

3. 调查问题的形式

调查问题形式上可以分为开放式问题、封闭式问题和半封闭式问题。①开放式问题：不给调查对象任何限制，由调查对象根据自身对问题的理解自由回答。优点是容易调动调查对象的积极性，从而获得较丰富的信息。缺点是容易离题，拒答率高，资料整理和分析比较麻烦。例如“您认为您的健康所面临的主要问题是什么？”、“您对当前医疗服务是怎样评价的？”。②封闭式问题：即将调查对象的回答限制在问题后面所列的答案中，常用的形式有“是与否”二项选择及多项选择。优点是问题便于理解和回答，应答率高，答案易于标准化和编码，整理分析起来比较方便。缺点是调查对象可以通过猜测和随便选答，从而影响调查的准确性。例如，“您近两周是否患过感冒？”回答选项是：A 是、B 否。又如，“您的文化程度是”，回答选项有：A 大学、B 专科、C 高中、D 初中、E 小学。③半封闭式问题：即将开放式与封闭式两者结合的问题，调查对象根据自身情况选择固定答案后，还需要做进一步文字表述。这种问题形式上比较灵活，不容易离题，而且提供的信息比封闭式问题稍多。例如，您对自己睡眠质量的评价是？回答选项是：A 好、B 不好（如选 B，具体的原因是_____）。

4. 问题设计原则

①用词简单直接，通俗易懂，避免采用专业术语。例如，“你所居住地是乙型脑炎的疫源地吗？”这里“疫源地”就是医学专业术语，一般人难以理解。②避免使用模糊和容易混淆的语句。问题和答案要避免含混不清的内容，否则不便获取准确的信息。如“您是否经常去医院看病？”，类似的用词还有“偶尔”、“有时”等，这些词的含义应在本次调查表填写说明中给以明确规定，如“经常”是指“每周一次或以上”。③避免双重问题或多重问题。双重和多重问题是指一个问题包含两个或多个内容，如“您是否抽烟喝酒？”这个问题包含了抽烟与喝酒两个内容，使调查对象难以回答。④避免诱导性提问。诱导性是指有意识地引导调查对象向某一方面回答。例如，“大多数人认为在公共场所吸烟是不道德的，您认为对吗？”。⑤敏感性问题处理。敏感性问题包括国家政策、伦理道德、生活行为、个人隐私等，调查对象面对敏感问题时比较谨慎，往往造成拒绝回答或虚假回答。问卷中这类问题应特别注意，其解决途径包括保密声明、匿名问卷等。此外，问题要精心设计，巧妙安排，常用的方法有对象转移法、假定法、随机应答技术等，例如，“对于艾滋病，有些人认为是行为不洁引起的，有些人不赞同此说法，您同意哪种观点？”以及“如果我国不实行计划生育政策，您愿意有几个孩子？”

5. 问题的排序原则

①符合逻辑。问题安排应符合人们的逻辑思维习惯，通常是先询问一般情况，如年龄、性别、文化程度、家庭情况，然后进入研究主体项目，如患病情况、治疗情况。②先一般后特殊。将一般性问题安排在前，特殊性问题安排在后。③先易后难。先安排容易回答的问题，难回答的问题在后。④先

熟悉后生疏。先安排熟悉的问题，生疏的问题在后。⑤妥善安排敏感性问题。一般将敏感性问题放在后面，但当这类问题较多时，可分散在其他问题中，以降低调查对象的敏感程度和拒答率。

6. 调查表评价

主要包括效度评价和信度评价。效度是指问卷所能反映调查对象真实情况的程度，它又称为有效性和准确性。效度可分为表面效度、内容效度和结构效度。信度是指在相同条件下对同一调查对象重复测量结果的一致程度，又称为精确度和稳定性，它主要反映了问卷调查结果中测量误差所占比例。测量信度的标准有重测信度、复本信度和分半信度等。好的调查表信度和效度均应该较高，其计算都可以通过 SAS、SPSS 等软件来实现。

调查表实例：某地区医务工作者对医疗保险态度的调查，调查表见表 11-14。

表 11-14 某地区医务工作者对医疗保险态度的调查表

医疗保险是一种个人储蓄积累和社会统筹相结合的保险，保险金分别由国家、单位和个人筹资组成，其中一部分作为个人账户基金由个人管理，另一部分作为社会统筹基金由医疗保险机构统一管理，充分发挥患病时个人自我救济和社会共同救济的作用。本问卷以不记名的形式进行有关医疗保险态度调查，每个问题有多个答案供选择。请您仔细阅读问题后，根据自己的真实情况，选择一个答案。

我们期待您真诚的合作，谢谢！

一、一般情况

年龄(岁)

☐

☐

性别

1 男

2 女

☐

☐

科室

1 内

2 外

3 妇

4 儿

5 行政

6 其他

☐

☐

☐

☐

☐

文化程度

1 中专

2 大专

3 本科

4 硕士

6 博士

7 其他

☐

☐

☐

☐

☐

☐

职称

1 高级

2 中级

3 初级

4 其他

☐

☐

☐

☐

职务

1 院级干部

2 中层干部

3 医生

4 护士

☐

☐

☐

☐

工作年限(年)

1 1~

2 5~

3 10~

4 15~

5 20~

6 25~7 30~

☐

☐

☐

☐

☐

☐

平均月收入(元)

1 <1500

2 1500~

3 3000~

4 4500~

5 6000~

☐

☐

☐

☐

☐

二、请表明您对下列医疗保险问题的态度

1. 您认为目前的医疗保险政策

1 很合理

2 较合理

3 不确定

4 较不合理

5 很不合理

☐

☐

☐

☐

☐

2. 您认为目前的医疗保险政策在体现国家福利、减轻职工后顾之忧方面的作用

1 很大

2 较大

3 不确定

4 较小

5 很小

☐

☐

☐

☐

☐

3. 您是否同意参与医疗保险有利于减轻单位负担

1 很同意

2 较同意

3 不确定

4 较不同意

5 很不同意

☐

☐

☐

☐

☐

4. 您认为实施医疗保险后，获益最大的是

1 国家

2 单位

3 个人

4 医院

5 均无

☐

☐

☐

☐

☐

5. 您是否愿意参加医疗保险

1 很愿意

2 较愿意

3 不确定

4 较愿意

5 很愿意

☐

☐

☐

☐

☐

填写日期

核查者签名

核查日期

11.5.5 调查资料信度与效度的评价及 SAS 实现

1. 信度的评价

信度指调查表测量结果的可靠性、稳定性和一致性，也即精确度。一般认为信度反映测量误差或观察误差引起的变异程度。答案的波动性越大，测量的信度就越低；答案的波动性越小，测量的一致性就越好，信度也就越高。评价信度常用的评价指标有重测信度、分半信度、内部一致性信度，常用的 Cronbach's α 系数即属于内部一致性测度。

(1) 重测信度是指用同一调查表在不同时间对同一组调查对象进行重复测量，两次结果间的一

致性即重测信度。它说明了使用同一测量工具重复测量时所得结果的稳定性。两次测量的时间间隔因调查种类不同而异,原则上应在调查的主要内容还未发生变化的期间内进行。重测信度是用 Pearson 相关公式计算的,读者可参阅有关专著。

(2) 分半信度是在一次测量后将条目分为相等但独立的两部分,分别计算两部分的得分,并以其相关系数作为信度指标。这实际上考察的是指标的一致性,因测量同一特征的指标间关系密切,故具有一致性则说明结果可信。两部分之间的相关系数(r_h)只表明一半条目的信度,需用 Spearman - Brown 公式来确定整个调查表的信度(r)。一般要求分半信度大于 0.7。

$$r = \frac{2r_h}{1 + r_h} \quad (11-1)$$

(3) 内部一致性信度是分半信度的推广。无须将条目分为两个部分,而是以条目之间的联系程度对信度做出估计。这里仅介绍克朗巴赫 α 系数(Cronbach's α 系数)。

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum S_i^2}{S^2} \right) \quad (11-2)$$

式中, K 为整个调查表的条目数, S_i^2 为第 i 个条目得分的方差, S^2 是整个调查表得分的方差。

克朗巴赫 α 系数取值在 0 到 1 之间, α 系数越高,信度越高,测量工具的内部一致性较好。

2. 效度的评价

效度即调查表的有效性和正确性,也称为准确度,意指调查表确实测定了研究者所要测定的特征以及测定的程度,测量指标或观察结果在多大程度上反映了所测对象信息的客观真实性。一个调查表的效度越高,说明调查的结果越能显示其所测对象的真正特征。效度考评可以从内容效度、结构效度、标准关联效度几个方面进行。

(1) 内容效度,是指调查表测量的内容与所要测量的内容之间的符合情况,即测定对象对问题的理解和回答是否与其希望询问的内容一致。常用的内容效度评价方法有两种。一是专家法,即请相关专家对测量工具题目和内容范围是否符合调查的目的进行分析,并作出判断;二是统计分析法,即从同一内容总体中抽取两套测量工具,分别对同一组答卷者进行测验,两套答卷的相关系数即可用来估计测量工具的效度。

(2) 结构效度,说明调查表的结构是否符合理论构想和框架,即调查表是否真正测量了所提出的理论构思。结构效度的评价较复杂,通常用多元统计分析中的探索性因子分析或证实性因子分析来揭示众多条目之间的内在联系,看是否与理论构想一致。

(3) 标准关联效度,是调查表所得数据与某种外部标准(校标)间的关联程度,常用测量数据与校标之间的相关系数表示。反映的是一个测量工具对个体将来的行为进行预测的准确性。外部标准可以是该调查表以外的一些客观指标或从另一种调查表得到的数据。

参 考 文 献

- [1] 方积乾. 卫生统计学(第 5 版). 北京: 人民卫生出版社, 2003. 32-224.
- [2] 孙振球. 医学统计学(第 2 版). 北京: 人民卫生出版社, 2008. 485-506.
- [3] 胡良平. 科研设计与统计分析. 北京: 军事医学科学出版社, 2012. 129-278.
- [4] 胡良平. SAS 统计分析教程. 北京: 电子工业出版社, 2010. 674-706.

第 12 章 制订试验设计方案的关键技术

12.1 随机化的种类和 SAS 实现方法

12.1.1 随机化的体现

随机化体现在如下三个方面。

(1) 抽样随机：每一个符合条件的受试对象参加试验的机会相同，即总体中每个个体有同等的机会被抽取进入样本之中。它保证所得到的样本具有代表性，使试验结论具有普遍意义。

(2) 分组随机：每个受试对象被分配到不同组（通常为对照组、不同处理组）中的机会相同。它保证大量难以控制的非处理因素在对比组间尽可能均衡一致，以提高各组间的可比性。

(3) 试验顺序随机：若每个受试对象需要先后接受多个处理，则需对处理顺序进行随机化，它使试验顺序对观测结果的影响在不同受试对象之间达到均衡。

表 12-1 是 20 例病人的基本信息，本节将以其为基础，阐述随机化的 SAS 实现。

表 12-1 20 例病人的基本信息

病人编号	1	2	3	4	5	6	7	8	9	10
性别	F	F	M	F	F	F	M	M	M	M
年龄	60	64	37	57	41	31	60	64	58	16
病人编号	11	12	13	14	15	16	17	18	19	20
性别	M	M	F	F	F	F	F	M	F	F
年龄	58	63	23	37	20	33	39	40	49	42

12.1.2 常用随机抽样方法的种类与 SAS 实现

目前在 SAS /STAT 模块中，有一个与调查设计有关的 surveyselect 过程。该过程可用来实现各种随机抽样，包括单纯随机抽样、系统随机抽样、分层抽样、无限随机抽样（有替换）、序贯随机抽样，以及按规模大小成比例概率抽样（PPS）等。

(1) 单纯随机抽样：就是从总体中不加任何分组、划类、排队等，完全随机地抽取受试对象。特点是每个个体（样品）被抽中的概率相等，各样品之间完全独立，彼此间无一定的关联性和排斥性。单纯随机抽样是其他各种抽样形式的基础。通常只是在总体样品之间差异程度较小和数目较少时，采用这种方法。

【例 12-1】 从 20 例病人中以简单随机化的方法抽取 10 名患者作为受试者。

【分析】 用 SAS 实现单纯随机抽样，程序名为 jcyx12_1.sas。

```
data jcyx12_1;
  input id sex $ age;
  cards;
  1    F    60
  2    F    64
  ...  ...  ...
  20   F    42
  ;
run;
```

```
ods html;
proc surveyselect data = jcyx12_1
  method=srs
    n=10 seed=10000 out=b;
run;
proc print data=b;
run;
ods html close;
```

【程序说明】 首先, 创建名为 jcyx12_1 的 SAS 数据集。然后, 调用 surveyselect 过程进行随机抽样, 并指定一些抽样选项, 其中“data =”指定输入数据集, 用于指定抽样框; “method =”用于指定随机抽样的方法, srs 表示单纯随机抽样; “n =”用于指定抽样大小(此处可以用“rate =”来替换, 用于指定抽样率); “seed =”用来指定随机种子数; “out =”用于指定输出数据集, 它包含所有被抽到的样本(可以添加选项“rep = number”来规定重复抽样的次数, 如“rep = 2”命令系统同时抽取 2 份样本)。最后用 print 过程把输出数据集 b 在输出窗口中显示出来。ods html 和 ods html close 命令用于把程序运行结果以网页格式保存下来。

【主要输出结果及解释】

obs	id	sex	age
1	1	F	60
2	4	F	57
3	7	M	60
4	11	M	58
5	13	F	23
6	14	F	37
7	17	F	39
8	18	M	40
9	19	F	49
10	20	F	42

“id”列即为抽取的患者编号。

(2) 系统随机抽样: 先按照某种顺序将总体中的个体(样品)从 $1 \sim N$ 相继编号, 并计算抽样距离 $K = N/n$ 。式中, N 为总体样品总数, n 为样本含量。然后在 $1 \sim K$ 中抽一随机数, i 作为样本的第一个被抽样品, 接着取 $i + K, i + 2K, \dots, i + (n - 1)K$, 直至抽够 n 个样品为止。值得注意的是, 当总体的被观察样品按顺序有周期趋势或有增加(减少)趋势时, 容易产生偏倚。

【例 12-2】 从 20 例病人中以系统随机抽样的方法抽取 5 名患者作为受试者。

【分析】 用 SAS 实现单纯随机抽样, 程序名为 jcyx12_2.sas。

```
ods html;
proc surveyselect data = jcyx12_1
method=sys
seed=10000 n=5 out=b;
control id;
run;

proc print data=b;
run;
ods html close;
```

【程序说明】 数据步骤与例 12-1 相同, 此处省略。surveyselect 过程与例 12-1 的不同之处在于选项 method = sys, 规定抽样方法为系统随机抽样法。此外, 还多了一个 control 语句, 该语句用于指定排序的变量。

【主要输出结果及解释】

obs	sex	age	id
1	M	37	3
2	M	60	7
3	M	58	11
4	F	20	15
5	F	49	19

“id”列即为抽取的患者编号。

(3)分层抽样：先依据一种或几种特征将总体分为若干个子总体，每个子总体被称为一个层，然后从每层中随机抽取一个子样本，这些子样本合起来就是所需要抽取的最终样本。分层抽样所得的样本具有较好的代表性，抽样误差较小。

【例 12-3】 以性别为分层因素，从 20 例病人中按比例抽取 10 名患者作为受试者。

【分析】 用 SAS 实现分层抽样，程序名为 jcyx12_3. sas。

```
ods html;
proc sort data=jcyx12_1;
  by sex;
run;
proc surveyselect data=jcyx12_1
  method=srs
  seed=10000 n=(6 4) out=b;
  strata sex;
run;

proc print data=b;
run;
ods html close;
```

【程序说明】 数据步骤与例 12-1 相同，此处省略。sort 过程用于将数据集 jcyx12_1 排序，by 语句指明按性别进行排序，此处排序方式为先 F 后 M。surveyselect 过程与例 12-1 的不同之处在于选项“n=(6 4)”，规定每层中的抽样大小，若每层抽样大小相同，可简写为“n=n₁”的形式。此外，还多了一个 strata 语句，该语句用于指定分层变量。

【主要输出结果及解释】

obs	sex	id	age	SelectionProb	SamplingWeight
1	F	1	60	0.5	2
2	F	4	57	0.5	2
3	F	6	31	0.5	2
4	F	15	20	0.5	2
5	F	17	39	0.5	2
6	F	20	42	0.5	2
7	M	9	58	0.5	2
8	M	11	58	0.5	2
9	M	12	63	0.5	2
10	M	18	40	0.5	2

“id”列即为抽取的患者编号。

(4)整群抽样：先将总体分成若干群，从中随机抽取几个群，抽中群内的全部个体(样品)组成样本。可将整群抽样转化为简单随机抽样。例如，从 10 个班随机抽取 5 个班作为受试对象，可看做从 10 个个体中随机抽取 5 个，即为简单随机抽样。

12.1.3 常用随机分组方法的种类与 SAS 实现

(1)完全随机化：直接对受试对象进行随机化分组，常通过掷硬币或随机数字表，或用计算机产生伪随机数来进行随机化，在事先或者实施过程中不作任何限制和干预或调整。分组后各组受试对象的例数不一定相等。

【例 12-4】 将 20 例患者完全随机地等分为两组。

【分析】 对 20 例患者进行编号，范围为 1~20。用 SAS 实现完全随机分组，程序名为 jcyx12_4. sas。

```

% let seed=10000;          /* 1 */
% let number=20;           /* 2 */
% let group1=group_1;      /* 3 */
% let group2=group_2;      /* 4 */
proc plan seed=&seed;       /* 5 */
  factors a=&number;
  output out=sheji;
run;
data b c;                  /* 6 */
  set sheji;
  subject=_n_;
  if a <= %sysevalf(&number/2)
  then do; output b; end;
  else do; output c; end;
run;
data d;                    /* 7 */
  set b;
  rename subject=&group1;
  drop a;
run;
data e;                    /* 8 */
  set c;
  rename subject=&group2;
  drop a;
run;
data f;                    /* 9 */
  merge d e;
run;
ods html;
proc print noobs;          /* 10 */
run;
ods html close;

```

【程序说明】 第 1 步至第 4 步为设置宏变量，依次为随机种子数、受试对象例数、第一组的名称和第二组的名称，使用时仅需根据实际情况修改这 5 个宏变量的取值即可，第 5 步是调用 plan 过程将数字进行随机化排列，结果输出至数据集 sheji 中。第 6 步对数据集 sheji 进行操作，产生患者编号，相当于依次将随机数分配给第 1 号到第 26 号患者，即为每个患者分配一个随机数字。随机数字小于等于 1/2 受试对象数的患者分入第一组，患者编号保存至数据集 b 中；其余患者分入第二组，患者编号保存至数据集 c 中。第 7 步、第 8 步分别对数据集 b 和 c 进行处理，主要是修改变量的名称，并分别保存在数据集 d 和 e 中。第 9 步对两个数据集 d 和 e 进行横向合并，得数据集 f。第 10 步打印输出数据集 f 中的内容。

【主要输出结果及解释】

group_1	group_2
3	1
5	2
7	4
9	6
10	8
12	11
13	15
14	18
16	19
17	20

“group_1”和“group_2”分别代表第 1 组和第 2 组，列中的数字代表分入该组患者的编号。

【例 12-5】 将 20 例患者完全随机地等分为四组。

【分析】 对 20 例患者进行编号，范围为 1~20。用 SAS 实现完全随机分组，程序名为 jcyx12_5.sas。

```
% let seed=10000;          /* 1 */
% let number=20;            /* 2 */
% let n_g=4;                /* 3 */
proc plan seed=&seed;        /* 4 */
  factors a=&number;
  output out=sheji;
run;
data d;                      /* 5 */
  input; cards;
run;
% macro multigroup;          /* 6 */
% do i=1 % to &n_g;
% let group&i=group_&i;
data b&i;
  set sheji;
  subject=_n_;
  if a <= % sysevalf (&number/&n_g
* &i) and
a >% sysevalf (&number/&n_g* (&i-1))
  then output;
run;
data c&i;
  set b&i;
  rename subject=&&group&i;
  drop a;
  run;
  data d;
  merge d c&i;
run;
% end;
% mend multigroup;
% multigroup;                /* 7 */
ods html;
proc print noobs;           /* 8 */
run;
ods html close;
```

【程序说明】 第1步至第3步为设置宏变量，依次为随机种子数、受试对象例数、分组数，使用时仅需根据实际情况修改这些宏变量的取值即可。第4步是调用 plan 过程将数字进行随机化排列，结果输出至数据集 sheji 中。第5步是建立空数据集，以便第6步编写的宏程序使用。第6步为建立宏程序 multigroup，作用是将受试对象进行随机等分，并自动为各组命名，命名规则为 group_组号。第7步是调用宏程序 multigroup。第8步打印输出最近产生的数据集中的内容。

【主要输出结果及解释】

group_1	group_2	group_3	group_4
5	3	1	2
7	10	11	4
9	13	18	6
12	14	19	8
16	17	20	15

group_1、group_2、group_3、group_4 分别代表第1组至第4组，列中的数字代表分入该组患者的编号。

(2) 分层随机化的实现：分层随机化，首先对可能影响试验过程和结果的主要混杂因素（如年龄、性别、病情、疾病分期等）进行分层，然后在每一层内进行完全随机化分组，最后分别合并为试验组和对照组。配对随机化和区组随机化可看成是分层随机化的实际应用。

【例 12-6】 以性别为区组因素，将表 12-1 中的 20 例患者随机分为两组。

【分析】 对 20 例患者进行编号，范围为 1~20。用 SAS 实现分层随机分组，程序名为 jcyx12_6.sas。

```
data b;                      /* 1 */
  set jcyx12_1;
  r=ranuni(10000);
run;
proc sort out=c;             /* 2 */
  by group;
run;
proc print noobs;            /* 6 */
proc sort out=f;            /* 5 */
  by group;
run;
ods html;
```

```

        by sex;
run;
proc rank data=c out=d;    /* 3 */
    ranks rank;
    var r;
    by sex;
run;
data e;                    /* 4 */
    set d;
    if sex = 'F' and rank <= 6 then
group=1;
    if sex = 'F' and rank > 6 then
group=2;
    if sex = 'M' and rank <= 4 then
group=1;
    if sex = 'M' and rank > 4 then
group=2;
run;
run;
ods html close;

```

【程序说明】 创建 jcyx12_1 数据集的数据步骤与例 12-1 相同, 此处省略。第 1 步为对数据集 jcyx12_1 中的每个患者产生一个随机数字 r, 并重新保存为数据集 b。第 2 步以性别变量 sex 为排序变量对数据集 b 进行排序, 并保存为数据集 c, 以便在后续调用的 rank 过程中使用 by 语句。第 3 步调用 rank 过程, 以性别变量 sex 为分组变量、以随机数字变量 r 为排序变量, 对患者进行编秩, 并保存为数据集 d。第 4 步根据性别变量 sex 和秩变量 rank 的取值, 使用条件语句将患者进行分组, 并保存为数据集 e。第 5 步, 根据分组变量 group 的取值, 对数据集 e 进行排序, 并保存为数据集 f。第 6 步打印输出最近产生的数据集中的内容。

【主要输出结果及解释】

id	sex	age	r	rank	group
1	F	60	0.62570	5	1
5	F	41	0.02578	1	1
15	F	20	0.23858	4	1
17	F	39	0.11789	3	1
19	F	49	0.64862	6	1
20	F	42	0.07519	2	1
3	M	37	0.24306	2	1
10	M	16	0.66552	4	1
11	M	58	0.18571	1	1
18	M	40	0.29631	3	1
2	F	64	0.88716	9	2
4	F	57	0.98645	12	2
6	F	31	0.77350	7	2
13	F	23	0.94311	11	2
14	F	37	0.84577	8	2
16	F	33	0.93964	10	2
7	M	60	0.79173	6	2
8	M	64	0.91723	7	2
9	M	58	0.94026	8	2
12	M	63	0.70521	5	2

“id”列为患者编号，“group”列为患者的分组结果。

(3)分层区组随机化：在多中心临床试验中，人们常以临床试验中心为分层因素，对受试对象进行分层随机化。

【例 12-7】 若有来自 2 个种系的大鼠，分别 24 只，每个种系分为各 3 个窝别，随机分为 4 组，每组 12 只。

【分析】 对 48 只大鼠进行编号，第 1 个种系的范围为 1~24，其中 1~8 为第一窝，9~16 为第二窝，17~24 为第三窝；第 2 个种系的范围为 25~48，其中 25~32 为第一窝，33~40 为第二窝，41~48 为第三窝。用 SAS 实现分层区组随机化，程序名为 jcyx12_7.sas。

```
proc plan seed=10000;          /* 1 */
  factors center = 2 ordered block
=3 ordered
  repeat =8;
  output out =a;
run;
data b;                        /* 2 */
  set a;
  subject =_n_;
  group =mod(repeat, 4) +1;
run;

proc sort out =c;              /* 3 */
  by group;
run;
ods html;
proc print noobs                /* 4 */
  var center block subject group;
run;
ods html close;
```

【程序说明】 第 1 步调用 plan 过程完成每个种系、每个窝别内大鼠的随机化排列。第 2 步实现大鼠的分组。第 3 步以分组变量 group 为排序变量对数据集 b 进行排序，并保存为数据集 c。第 4 步打印输出最近产生的数据集中的内容。

【主要输出结果及解释】

center	block	subject	group
1	1	2	1
1	1	3	1
1	2	9	1
1	2	11	1
1	3	18	1
1	3	23	1
2	1	25	1
2	1	27	1
2	2	38	1
2	2	40	1
2	3	43	1
2	3	48	1
1	1	5	2
1	1	7	2
1	2	13	2
1	2	14	2
1	3	17	2
1	3	24	2
2	1	26	2
2	1	31	2
2	2	36	2

2	2	37	2
2	3	46	2
2	3	47	2
1	1	1	3
1	1	4	3
1	2	10	3
1	2	16	3
1	3	19	3
1	3	20	3
2	1	29	3
2	1	30	3
2	2	33	3
2	2	35	3
2	3	42	3
2	3	44	3
1	1	6	4
1	1	8	4
1	2	12	4
1	2	15	4
1	3	21	4
1	3	22	4
2	1	28	4
2	1	32	4
2	2	34	4
2	2	39	4
2	3	41	4
2	3	45	4

“center”列为种系，“block”为窝别，“subject”列为大鼠编号，“group”列为大鼠的分组结果。

12.2 样本含量和检验效能估计的 SAS 实现

当假设检验的结果为“阴性”，即不拒绝原假设时，很少有研究者去分析假设检验的检验效能，进而判断“阴性”结果的可信度如何。造成这种状况的原因，一是研究者没有认识到样本含量估计与检验效能分析对科学研究的重要性，二是研究者不知道如何进行样本含量估计与检验效能分析。为此，本节将为读者介绍常见情况下样本含量估计与检验效能分析的具体方法，并给出相应的 SAS 程序。

12.2.1 估计样本含量的意义

样本含量，亦即观察例数的多少，在数理统计上习惯于称为样本大小。在医学科研工作中，无论抽样调查，临床试验或试验研究，正确确定样本含量都是一项十分重要的内容，也是一个需要科研人员给予认真考虑的问题。例如，要比较某种中药与某种西药对甲型 H1N1 流感患者的治疗效果，各组需要观察多少患者？根据假设检验的原理，若样本太小，会使本来存在的差别不能显示出来，难以获得正确的研究结果，结论也缺乏充分的依据；反之，若样本太大，也会增加实际工作中的困难，不必要地浪费人力、物力、财力和时间，此外，由于样本太大，可能投入不足，科研过程的

质量控制下降,从而引入不必要的干扰因素,对研究结果也可能造成不良影响。

事实上,在一个科研项目中,所需样本含量的大小与许多因素有关,涉及统计研究设计的类型(调查设计、实验设计和临床试验设计),对结果精确度要求的高低和研究进度的快慢以及资料的性质和研究目的,需借助适当的公式,进行样本含量的估计。研究者可以根据需要和可能来确定一个适合的样本含量。另一方面,当科研人员或医务工作者阅读专业文献时,对于那些假设检验的阴性结果($P>0.05$),有必要复核样本含量和检验效能(若对比组的参数之间确实存在有意义的差别,假设检验能够发现这种差别的能力,被称为检验效能或把握度)是否偏低,以便正确分析假设检验的结论。

12.2.2 确定样本含量时应具备的条件

要想确定一个研究项目所需要的样本含量,凭空想象是没有科学依据的,必需提供一些起码的前提条件,才能作出有理有据的估算。估计样本含量时,通常需要提供的前提条件有如下几项。

(1)定出检验水准或显著性水平,即事先规定本次试验允许犯 I 型(或假阳性)错误的概率 α ,通常规定 $\alpha = 0.05$,同时还应明确是单侧检验还是双侧检验。 α 定得越小,所需的样本含量越大。同一问题,在其他条件不变的情况下,用单侧检验比用双侧检验所需的样本含量要少,但究竟该用单侧检验还是双侧检验取决于专业知识,而不是人的意愿。

(2)提出所期望的检验效能或把握度 $\text{power} = 1 - \beta$ (这里的 β 为允许犯 II 型或假阴性错误的概率),即在特定的 α 水准下,若总体对比的参数之间确实存在着差别,此时该次试验能发现此差别的概率。要求的检验效能越大,所需的样本含量就越大,实际上,检验效能由犯 II 型错误的概率 β 的大小所决定。在科研设计时,检验效能不宜低于 0.75,否则检验的结果很可能反映不出总体的真实差别,出现非真实的阴性结果。

(3)必须知道由样本推断总体的一些信息。比较两总体均数或概率的差别时,应当知道总体参数间的差值 δ 的信息。例如,两总体均数间的差值 $\delta = \mu_1 - \mu_2$ 的信息,两总体概率间的差值 $\delta = \pi_1 - \pi_2$ 的信息。有时研究者很难得到总体参数的信息,但可以用专业上(临床上)认为有意义的差值代替,如平均舒张期血压的差值 $\geq 0.69\text{kPa}$ (5mmHg),白细胞的平均差值为 $0.5 \times 10^9/\text{L}$ (500 个/ mm^3) 等,也有人主张用 0.25 或 0.50 的标准差估计总体均数间的差值。当然也可以根据试验的目的人为规定,如规定试验的新药有效概率必须超过标准药物有效概率的 30% 才有推广价值等。此外,确定两均数比较的样本含量还需要有关总体标准差 σ 的信息。这些信息也可以通过查阅资料,借鉴前人的经验或者进行预试验来寻找参考值。

(4)其他方面。如研究类型、设计类型、比较类型和主要评价指标的性质等。

12.2.3 估计总体均值时样本含量估计的 SAS 实现

【例 12-8】 某临床医师欲了解高脂血症患者的血液中三酰甘油的含量,现已知正常人血液中三酰甘油含量的标准差是 0.35,这次希望绝对误差不超过 0.2,取 $\alpha = 0.05$,在以上条件下,要估计患者的平均三酰甘油含量,需要调查多少患者?

【分析】 该研究是一个单组设计定量资料置信区间的估计问题。已知样本标准差、置信区间的显著性水平以及测量结果的绝对误差,欲估计所需调查的样本量,可借助下面的 SAS 程序进行计算,SAS 程序名为 jcyx12_8.sas。

```
proc power;
  onesamplemeans ci = t sides = 2
  alpha = 0.05 halfwidth = 0.2
  stddev = 0.35
  ntotal = . probwidth = 0.3;
run;
```

【程序说明】 选项“ci=t”表示要估计总体均数的基于 t 分布的置信区间，该语句也可以简写为“ci”；选项“sides=2”（默认设置）中的 2 表示要求双侧对称的置信区间；需要注意的是，选择双侧还是单侧，不是随意确定的，而是以专业知识为依据，如果能够确定该指标不会高于或低于某一个值，则选择单侧。选项“alpha=0.05”（默认设置）中的 0.05 表示要求置信度为 $100(1-0.05)\%$ 的置信区间；选项“halfwidth=0.2”中的 0.2 为误差，即置信区间的半宽度；选项“stddev=0.35”中的 0.35 为已知的标准差；选项“probwidth=0.3”中的 0.3 为获得半宽不超过最大误差值 0.2 的置信区间的概率，此值定得越大，所需要的样本量也会越大，如取“probwidth=0.90”，则 $n=20$ 。在应用时，读者只要根据实际情况修改这些关键字所代表的变量的取值，即可得到结果。在实际使用中，为保险起见，“probwidth”的值应取大一些，如 0.95 或 0.99。

【主要输出结果及解释】

Computed N Total	
Actual Prob(Width)	N Total
0.349	12

结果：需调查大约 12 人。

12.2.4 估计总体率时样本含量估计的 SAS 实现

1. 当 π 或 P 接近 0 或 1 时

【例 12-9】 某地区现调查 HBsAg 阳性率，过去调查的结果为 $P=10\%$ ，本次调查容许误差不超过 $0.1P$ ， $\alpha=0.05$ （双侧），估计应调查人数。

【分析】 该研究是一个结果变量为二值变量的单组设计定性资料阳性率估计问题，已知阳性率的估计值、阳性率的调查容许误差（即阳性率置信区间的半宽）及显著性水平，欲估计所需调查人数。由于 $P=10\%$ ，比较接近于 0，故需按下面的 SAS 程序进行样本含量估计（当 P 比较接近于 1 时也应如此），程序名为 jcyx12_9.sas。

```
% let P=0.10;
% let delta=0.01;
% let alpha=0.05;
data jcyx12_9;
    n = (probit (1 - &alpha/2)/arsin
    (&delta
    /sqrt (&P* (1 - &P))))** 2;
    n1 = ceil(n);

file print;
put #3 @ 15 '需调查大约' n1 '人。';
run;
```

【程序说明】 $P=0.10$ 中的 0.10 为阳性率， $\delta=0.01$ 中的 0.01 是误差的限值， $\alpha=0.05$ 中的 0.05 为可能犯第一类错误的概率，上述数字需要根据实际情况修改。

【主要输出结果及解释】

结果：需调查大约 3457 人。

2. 当 π 或 P 接近 0.5 时

【例 12-10】 拟用抽样调查了解某地小学生蛔虫感染率。假定以往该地小学生蛔虫感染率 $P=50\%$ ，要求误差不超过 3%，如取 $\alpha=0.05$ ，问需调查多少人？

【分析】 该研究是一个结果变量为二值变量的单组设计定性资料感染率估计问题，已知感染率的估计值、感染率的调查容许误差（即感染率置信区间的半宽）及显著性水平，欲估计所需调查

人数。由于 $P=50\%$ ，可按下面的 SAS 程序进行样本含量估计 (P 比较接近于 50% 时均应如此；当 P 未知时，按 $P=0.50$ 计算)，程序名为 jcyx12_10. sas。

```
% let P=0.5;
% let delta=0.03;
% let alpha=0.05;
data jcyx12_10;
  n = (probit ((1 - &alpha/2)) /
    &delta)** 2* &p*
  (1 - &p);
  n1=ceil(n);
  file print;
  put #3 @ 15 '需调查大约' n1 '人。';
run;
```

【程序说明】“ $p=0.5$ ”中的 0.5 为感染率，“ $\text{delta}=0.03$ ”中的 0.03 是误差的限值，“ $\text{alpha}=0.05$ ”中的 0.05 为可能犯第一类错误的概率，上述数字需要根据实际情况修改。

【主要输出结果及解释】

结果：需调查大约 1068 人。

12.2.5 单组设计均值与率的检验时样本含量估计的 SAS 实现

1. 均值的比较

【例 12-11】某研究者欲研究再生障碍性贫血患者与正常人的外周血中单核细胞血小板内皮细胞黏附因子水平有无差异，选择 8 例再生障碍性贫血患者进行预试验，测得这些患者细胞黏附因子的平均水平为 0.40，标准差为 0.20，已知正常人的外周血中单核细胞血小板内皮细胞黏附因子的平均水平为 0.26。欲进行进一步的研究，取 $\alpha=0.05$ 、 $\beta=0.05$ ，试问需要多少患者参与研究才可得出两类人群外周血中单核细胞血小板内皮细胞黏附因子水平不同的结论？

【分析】本资料为单组设计定量资料，因专业上暂未认定再生障碍性贫血患者与正常人体内细胞黏附因子的平均水平存在必然的大小关系，所以应选择双侧检验。其中， $\delta=0.40-0.26=0.14$ 、 $\sigma=0.20$ 、 $\alpha=0.05$ 和 $\beta=0.05$ ，借助 SAS 软件的 power 过程进行分析，程序如下，程序名为 jcyx12_11. sas。

```
proc power;
  onesamplemeans test=t sides=2
  mean=0.14 stddev=0.20
  ntotal=. power=0.95;
run;
```

【程序说明】选项“test=t”表示对均值进行 t 检验；sides=2 或者省略此选项表示进行双侧检验，若 sides=1 则表示进行单侧检验；mean=0.14 表示两总体均值差值的估计值为 0.14；stddev=0.20 表示总体标准差的估计值为 0.20；ntotal=. (缺失值) 表示需要估计总的样本含量；power=0.95 表示检验效能的值为 0.95。程序中的数字可根据实际情况进行修改。

【主要输出结果及解释】

Computed N Total	
Actual Power	N Total
0.953	29

可见，需要 29 例再生障碍性贫血患者参与研究，才能使得把握度达到 95%。

2. 率的比较

【例 12-12】教育部目前就《通用规范汉字表》(征求意见稿)公开征求意见，“琴”、“亲”、“魅”等 44 个汉字的字形拟进行调整，此外，还恢复了 51 个异体字。但新浪网的调查结果显示，

90.3%的网友表示反对。请问：需询问多少网友才能有 95% 的把握得出持反对意见的比例超过 90% 的结论？取 $\alpha = 0.05$ 。

【分析】这是一个结果变量为定性变量的单组设计资料的样本含量估计问题。已知检验效能、样本率和总体率的估计值及显著性水平，欲估计达到检验效能的最小样本含量。所需的 SAS 程序如下，程序名为 jcyx12_12.sas。

```
% let p=0.903;
% let p0=0.900;
% let delta = 0.003; /* delta = p -
p0; */
% let alpha=0.05;
% let beta=0.05;
data jcyx12_12;
    n = (((probit(1 - &alpha)) * sqrt
(&p0 * (1 - &p0))
    + (probit(1 - &beta)) * sqrt(&p *
(1 - &p))) /
    &delta)** 2;

n1 = ceil(n);
file print;
put #3 @ 15 '需调查大约' n1 '名网
友。';
run;
```

【程序说明】5 个“%let”语句分别用于设置样本率、总体率、误差限值、允许犯第一类错误的概率和允许犯第二类错误的概率。对于类似的问题，只需要根据实际情况对 5 个“%let”语句中的参数进行修改即可。

【主要输出结果及解释】

结果：需调查大约 106769 名网友。

12.2.6 配对设计均值与率的检验时样本含量估计的 SAS 实现

1. 均值的检验

此情形下的样本含量估计可先将配对设计转化为单组设计(用差值作为结果变量，差值的均值与“0”进行比较)，按单组设计的方法进行处理。

【例 12-13】某医生试验某种增加白细胞的药物的疗效，选择 9 例患者进行预试验，患者用药前后白细胞数量差值的标准差为 $2.5 \times 10^9/L$ 。下一步欲进行正式的临床试验，取 $\alpha = 0.05$ 、 $\beta = 0.10$ ，且专业上认为白细胞平均增加 $1.0 \times 10^9/L$ 才算该药临床有效，试问需要多少患者参与试验才可得出该药有临床效果的结论？

【分析】本资料为自身配对设计定量资料，因专业上认为白细胞平均增加 $1.0 \times 10^9/L$ 才算该药临床有效，所以应选择单侧检验。 $\delta_d = 1.0 \times 10^9$ 、 $\sigma_d = 2.5 \times 10^9$ 、 $\alpha = 0.05$ 、 $\beta = 0.10$ ，借助 SAS 软件的 power 过程进行分析，程序如下，程序名为 jcyx12_13.sas。

```
proc power;
    onesamplemeans test=t sides=1
    mean=1.0 stddev=2.5

ntotal=. power=0.9;
run;
```

【程序说明】选项“test=t”表示对均值进行 t 检验；“sides=1”表示进行单侧检验，若“sides=2”或者省略此选项则表示进行双侧检验；“mean=1.0”表示治疗前后成对数据差值的均值的估计值为 1.0；“stddev=2.5”表示治疗前后成对数据差值的标准差的估计值为 2.5；ntotal=. (缺失值) 表示需要估计总的样本含量；power=0.90 表示检验效能的值为 0.90。程序中的数字需要根据实际情况进行修改。

【主要输出结果及解释】

Computed N Total	
Actual Power	N Total
0.900	55

可见，需要 55 例患者进入正式临床试验，才能使得把握度达到 90%。

2. 率的检验

【例 12-14】 某菌种接种于甲、乙两种培养基的结果如下：甲阳性乙阴性的概率为 $\pi_{+-} = 0.04$ ，甲阴性乙阳性的概率为 $\pi_{-+} = 0.24$ ，设 $\alpha = 0.05$ （双侧）， $\beta = 0.10$ ，问应该用多少样本对子数？

【分析】 这是一个结果变量为二值变量的配对设计试验样本含量估计问题。已知两种方法的两个不一致率及假设检验中容许犯 I、II 型错误的概率，欲估计最低样本含量。所需的 SAS 程序如下，程序名为 jcyx12_14.sas。

<pre>proc power; pairedfreq dist =NORMAL discproptdiff =0.20 totalproptdisc =0.28 method =CONNOR npairs =. power =0.90; run;</pre>	<pre>proc power; pairedfreq dist =NORMAL discproptdiff =0.20 totalproptdisc =0.28 method =MIETTINEN npairs =. power =0.90; run;</pre>
--	---

【程序说明】 程序中的选项“dist = NORMAL”定义检验统计量的分布类型；选项“discproptdiff = 0.20”中的 0.20 表示两个不一致概率 π_{-+} （=0.24）与 π_{+-} （=0.04）的差值；选项“totalproptdisc = 0.28”中的 0.28 表示两个不一致概率 π_{-+} （=0.24）与 π_{+-} （=0.04）的和；选项“method = CONNOR”和“method = MIETTINEN”定义两种不同的正态近似算法。程序中的数字需要根据实际情况修改。

【主要输出结果及解释】

结果：所需样本对数为 70（Connor 近似法）或 61（Miettinen 近似法）。

【例 12-15】 某眼科研究，患者一只眼采用某种治疗视力减低的新疗法，另一只眼采用传统疗法。传统疗法的失败率为 40%，新疗法预计可使失败率降至 20%，假定两法疗效不一致的患者数占总患者数的比例 $p = 30\%$ ，取 $\alpha = 0.05$ （双侧）， $\beta = 0.10$ ，问试验所需患者数是多少？

【分析】 这也是一个结果变量为二值变量的配对设计研究样本含量估计问题。已知两种方法各自的治疗失败率、两法疗效不一致的患者数占总患者数的比例及假设检验中容许犯 I、II 型错误的概率，欲估计最低样本含量。本例所知道的前提条件与例 12-14 不一样，故不能按前例的方法进行样本含量估计，但可以按下面的算法进行估计。所需的 SAS 程序如下，程序名为 jcyx12_15.sas。

<pre>% let p0=0.40; % let p1=0.20; % let p=0.30; % let alpha=0.05; % let beta=0.10; data jcyx12_15; n = (probit(1 - &alpha/2) +probit (1 - &beta)) ** 2 * &p/ (&p0 - &p1) ** 2;</pre>	<pre>n1 =ceil (n); file print; put #3 @ 15 '试验所需患者数为' n1 ' 例。'; run;</pre>
---	--

【程序说明】 程序中 5 个“% let”语句中的参数依次为传统方法治疗失败率、新疗法的治疗失败率、两法疗效不一致的患者数占总患者数的比例及假设检验中容许犯 I、II 型错误的概率。对于类似的问题，只需修改本程序前面 5 个“% let”语句中的参数即可。

【主要输出结果及解释】

结果：试验所需患者数为 79 例。

12.2.7 成组设计均值与率的差异性检验时样本含量估计的 SAS 实现

1. 均值的检验

【例 12-16】 在动物镇咳试验中，比较中药复方 I 与复方 II 使小鼠推迟发生咳嗽的时间，复方 I 与复方 II 的平均数分别为 31.67s 和 44.00s (即 $\delta = 44.00 - 31.67 = 12.33$ s)。设两组标准差相等，且为 25s, $\alpha = 0.05$ (双侧), $\beta = 0.10$, 要得出两组之间的差别有统计学意义的结论，问需要用多少只小鼠？

【分析】 这是一个结果变量为定量变量的成组设计研究的样本含量估计问题。已知两组的样本均值、样本标准差及假设检验中容许犯 I、II 型错误的概率，假定两组的样本含量相等，欲估计最低样本含量。所需的 SAS 程序如下，程序名为 jcyx12_16.sas。

```
proc power;
  twosamplemeans test=diff sides=2
  meandiff =12.33 stddev =25.00
  groupweights = (1 1)
  power =0.90
  ntotal = .
  ;
  run;
```

【程序说明】 程序中的选项“test = diff”表示对均差进行 t 检验；“meandiff =12.33”中的 12.33 为两均值的差值；“groupweights = (1 1)”表示两组样本含量的分配比为 $n_1:n_2 = 1:1$ 。程序中的数字需要根据实际情况修改。

【主要输出结果及解释】

Computed N Total	
Actual Power	N Total
0.902	176

可见，两组共需 176 只小鼠，每组需要 88 只小鼠。

2. 率的检验

【例 12-17】 拟研究两种抗菌药物 (其中一种为对照药) 对某感染性疾病的治疗效果，经预试验，试验药有效率为 80%，对照药有效率为 60%，今要做正式临床试验，取 $\alpha = 0.05$ (双侧), $\beta = 0.10$, 且要求试验组与对照组的样本含量之比为 2:1, 问每组需要观察多少例患者？

【分析】 这是一个结果变量为定性变量的成组设计研究的样本含量估计问题。欲估计最低样本含量，所需的 SAS 程序如下，程序名为 jcyx12_17.sas。

```
proc power;
  twosamplefreq test=PChi
  groupproportions = (0.80 0.60)
  sides=2 groupweights = (2 1)
  alpha=0.05 power =0.90
  ntotal = .;
  run;
```

【程序说明】 选项“test = PChi”说明处理数据时采用 Pearson χ^2 检验；选项“groupprop-

ortions = (0.80 0.60)”中的 0.80 和 0.60 分别是两组的有效率,“groupweights = (2 1)”说明两组样本量为 2:1;其他选项的说明同前述各例。程序中的数字需要根据实际情况修改。

【主要输出结果及解释】

Computed N Total	
Actual Power	N Total
0.901	240

可见,两组共需 240 例,其中试验组 160 例,对照组 80 例。

12.2.8 单因素多水平设计均值与率的检验时样本含量估计的 SAS 实现

1. 均值的检验

【例 12-18】 某研究者欲研究两种含电解质饮料(饮料甲、饮料乙)在降低长跑运动员体内乳酸形成中的效果,以纯净水为对照,在 10 公里的路程中,为每名运动员始终提供三种解渴物质之一,长跑结束后,测量每名运动员体内的乳酸含量。小规模预实验结果表明,长跑结束后,始终饮用纯净水的运动员体内的乳酸含量为 35.60,始终饮用饮料甲的运动员体内的乳酸含量为 30.20,始终饮用饮料乙的运动员体内的乳酸含量为 25.90,各组的标准差均为 3.75。欲进行进一步的研究,取 $\alpha=0.05$ 、 $\beta=0.10$,且要求饮用纯净水、饮料甲、饮料乙的运动员比例为 2:1:1,并进行三组间均数的两两比较,试问至少需要多少运动员参与研究才较为充分,各组分别需要多少运动员?

【分析】 本资料为单因素三水平设计定量资料,研究者考虑进行多组间均数的两两比较。已知 $\mu_1=35.60$ 、 $\mu_2=30.20$ 、 $\mu_3=25.90$ 、 $\sigma=3.75$ 、 $\alpha=0.05$ 和 $\beta=0.10$,各组间样本量之比为 2:1:1,借助 SAS 软件的 power 过程来估算所需的样本含量,程序如下,程序名为 jcyx12_18.sas。

```
proc power;
  onewayanova
  groupmeans = 35.60 |30.20 |25.90
  stddev = 3.75 groupweights = (2 1 1)
  alpha = 0.05 power = 0.90
  contrast = (1 -1 0) (1 0 -1) (0 1 -1)
  ntotal = .;
run;
```

【程序说明】 选项“groupmeans = 35.60 | 30.20 | 25.90”表示 3 个总体均值的估计值分别为 35.60、30.20、25.90;“stddev = 3.75”表示各总体标准差的估计值为 3.75;“groupweights = (2 1 1)”表示与均值对应的各组间样本例数之比为 2:1:1;“alpha = 0.05”表示一类错误的概率为 0.05;ntotal = . (缺失值)表示需要估计总的样本含量;“power = 0.90”表示检验效能的值为 0.90;“contrast = (1 -1 0) (1 0 -1) (0 1 -1)”依次表示对第一组与第二组、第一组与第三组、第二组与第三组的均值进行比较。程序中的数字需要根据实际情况进行修改。

【主要输出结果及解释】

		Computed N Total			
Index	Contrast	Actual Power	N Total		
1	1 -1 0	0.928	36		
2	1 0 -1	0.901	12		
3	0 1 -1	0.909	68		

可见,3 次组间均值的两两比较,最多需要的样本含量为 68。即需要 68 例高血压患者参与研究,才能使得各组间均数的两两比较把握度都达到 90% 以上。由于各组间样本比例为 2:1:1,所以纯净水组样本量为 34 例,饮料甲和饮料乙组的样本量均为 17 例。

2. 率的检验

【例 12-19】有三种驱肠道蠕虫寄生虫药，进行预试验，初估甲药经粪便检查，虫卵阴转频率甲药为 80%，乙药为 85%，丙药为 95%，设 $\alpha=0.05$ 、 $\beta=0.10$ ，问各药经正式临床试验需试验多少病人？

【分析】这是一个结果变量为二值变量的单因素三水平设计试验研究的样本含量估计问题。所需的 SAS 程序如下，程序名为 jcyx12_19.sas。

```
% let alpha=0.05;
% let beta=0.10;
% let p=0.80, 0.85, 0.95;
% let lam=12.65;
data jcyx12_19;
  p1=max(&p);
  p2=min(&p);

  n=ceil(&lam/(2*(arsin(sqrt(p1))-
    arsin(sqrt(p2)))**2));
  file print;
  put #3 @10 '每组所需样本量为' n '
  例。';
run;
```

【程序说明】“% let”语句用来设定各参数的取值。“alpha = 0.05”中的 0.05 为可能犯第一类错误的概率，“beta = 0.10”中的 0.10 为可能犯第二类错误的概率，“p = 0.80, 0.85, 0.95”中的 0.80, 0.85, 0.95 是各组的频率，修改时把各组的频率都写在这里，不同频率之间用逗号分隔，“lam = 12.65”中的 12.65 通过查 λ 值表得出，上述数字需要根据实际情况修改。

【主要输出结果及解释】

结果：每组所需样本量为 112 例。

12.2.9 检验效能分析概述

检验效能由Ⅱ型错误概率 β 的大小所决定，即 $\text{power} = 1 - \beta$ 。当假设检验出现“阴性”结果（ $P > 0.05$ ）时，有必要复核样本含量和检验效能是否偏低，以便正确分析假设检验“阴性”结论的正确性。当假设检验为“阴性”结论（ $P > 0.05$ ）时，若检验效能偏低，说明试验的样本含量不够，应进一步增大样本含量，继续进行试验。若检验效能足够大，可下“接受原假设”之阴性结论，尽管此时仍可能犯Ⅱ类错误，但所犯Ⅱ类错误的概率 β 在可以接受的范围之内。

检验效能分析与样本含量估计通常互为同一个问题的两个方面。因此，同类研究的检验效能分析与样本含量估计的计算公式通常也是可以互相推导的。

12.2.10 单组、配对设计定量资料假设检验时检验效能的 SAS 实现

【例 12-20】研究者测量了 25 例再生障碍性贫血患者外周血中单核细胞血小板内皮细胞黏附因子的水平，得这些患者细胞黏附因子的平均水平为 0.40，标准差为 0.20。已知正常人的外周血中单核细胞血小板内皮细胞黏附因子的平均水平为 0.26，欲比较两类人群外周血中单核细胞血小板内皮细胞黏附因子水平是否相同，取 $\alpha=0.05$ ，试计算其检验效能。

【分析】本资料为单组设计定量资料，因专业上暂未认定再生障碍性贫血患者与正常人体内细胞黏附因子的平均水平存在必然的大小关系，所以应选择双侧检验。已知 $\delta = 0.40 - 0.26 = 0.14$ 、 $\sigma = 0.20$ 、 $\alpha = 0.05$ 和 $n = 25$ ，借助 SAS 软件的 power 过程进行分析，程序如下，程序名为 jcyx12_20.sas。

```
proc power;
  onesamplemeans test=t sides=2
  mean=0.14 stddev=0.20

  ntotal=25 power=.;
run;
```

【程序说明】 选项“test = t”表示对均值进行 t 检验；“sides = 2”表示进行双侧检验；“mean = 0.14”表示两总体均值差值的估计值为 0.14；“stddev = 0.20”表示总体标准差的估计值为 0.20；“ntotal = 25”表示总的样本含量为 25；power = . (缺失值) 表示需要计算检验效能的值。程序中的数字需要根据实际情况进行修改。

【主要输出结果及解释】

Computed Power
Power
0.919

可见，样本含量为 25 时，检验效能为 0.919。

【例 12-21】 某医生选择 40 例患者进行试验，测得患者用药前后白细胞数量差值的均值为 $1.0 \times 10^9/L$ ，标准差为 $2.5 \times 10^9/L$ 。欲比较患者用药后白细胞数量是否确实升高，取 $\alpha = 0.05$ ，试计算其检验效能。

【分析】 本资料为自身配对设计定量资料，欲比较患者用药后白细胞数量是否确实高于用药前的数量，所以应选择单侧检验。已知 $\delta_d = 1.0$ 、 $\sigma_d = 2.5$ 、 $\alpha = 0.05$ 、 $n = 40$ ，借助 SAS 软件的 POWER 过程进行分析，程序如下，程序名为 jcyx12_21.sas。

```
proc power;                                ntotal=40 power = .;
  onesamplemeans test=t sides=1           run;
  mean=1.0 stddev=2.5
```

【程序说明】 选项“test = t”表示对均值进行 t 检验；“sides = 1”表示进行单侧检验；“mean = 1.0”表示两总体均值差值的估计值为 1.0；“stddev = 2.5”表示总体标准差的估计值为 2.5；“ntotal = 40”表示总的样本含量为 40；power = . (缺失值) 表示需要估计检验效能的值。程序中的数字需要根据实际情况进行修改。

【主要输出结果及解释】

Computed Power
Power
0.800

可见，样本含量为 40 时，检验效能为 0.800。

12.2.11 成组设计均值与率的差异性检验时检验效能的 SAS 实现

1. 均值的检验

【例 12-22】 某新降压药(试验药)与某老降压药(对照药)进行临床试验，每组各有 30 例受试对象，试验组用药后舒张压下降 1.5 ± 0.83 (kPa)，对照组用药后舒张压下降 1.5 ± 0.83 (kPa)。试计算此统计分析的检验效能。

【分析】 这是一个成组设计定量资料假设检验的检验效能分析问题。根据已知条件，可用下面的 SAS 程序进行检验效能分析，程序名为 jcyx12_22.sas。

```
proc power;                                groupns = (30 30)
  twosamplemeans test=diff                power = .;
  groupmeans = 1.50 | 1.20                run;
  stddev = 0.83
```

【程序说明】 选项“groupmeans = 1.50|1.20”中的 1.50 和 1.20 分别为试验组和对照组的均值，选项“stddev = 0.83”中的 0.83 是两组的混合标准差，选项“groupns = (30 30)”中的 30 和 30 分别为两组的例数。上述数字需要根据实际情况修改。

【主要输出结果及解释】

```
Computed Power
Power
0.280
```

可见，检验效能(Power) = 0.280 < 0.75，检验效能不够充分。

2. 率的检验

【例 12-23】 某国产新药(试验药)需对某病种进行临床试验，以考核其疗效，用同类进口药作对照，试验组 50 例患者，有效率为 70.0%，对照组 45 例患者，有效率为 84.4%，经 χ^2 检验无统计学意义($P > 0.05$)，试计算其检验效能。

【分析】 这是一个成组设计定性资料假设检验的检验效能分析问题。根据已知条件，可用下面的 SAS 程序进行检验效能分析，程序名为 jcyx12_23.sas。

```
proc power;
  twosamplefreq test = PChi
  groupproportions = (0.700 0.844)
  groupns = 50 | 45 sides = 2
  alpha = 0.05
  power = .;
run;
```

【程序说明】 选项“groupproportions = (0.700 0.844)”中的 0.700 和 0.844 分别是两组的有效率；“groupns = 50|45”指明试验组和对照组的样本量分别为 50 例和 45 例；选项“sides = 2”中的 2 表示双侧检验(为默认状态，可省略)；选项“alpha = 0.05”中的 0.05 表示犯第 I 类(或假阳性)错误的概率(为默认状态，可省略)。上述数字需要根据实际情况修改。

【主要输出结果及解释】

```
Computed Power
Power
0.379
```

可见，检验效能(Power) = 0.379 < 0.75，检验效能不够充分。

12.3 多因素试验设计类型的构造及辨析

所谓试验设计类型，就是安排因素及其水平组合的一种结构，常需体现以下情况：即因素及水平的数量、因素的性质、因素的水平之间的组合关系、因素与受试对象之间的关系、因素是否同时施加、因素对结果的影响是否地位平等。正确识别了试验设计类型，有利于正确选择统计分析方法处理实际资料。本节将介绍生物医学与临床研究中最为适用的几种试验设计类型。

12.3.1 随机区组设计和具有一个重复测量的单因素设计

1. 随机区组设计

实际问题：怎样才能尽可能消除来自受试对象的重要非试验因素对结果的影响，使两种以上的处理所产生的效应能更真实地表现出来？换句话说，如何才能把三种配对形式拓展或镶嵌到单因素多水平设计中去？下面以几个实例进行演示。

(1)如何才能把条件相近者配对设计拓展或镶嵌到单因素多水平设计中?

【例 12-24】 某研究者欲比较 3 种抗癌药物对小白鼠肉瘤的抑瘤疗效,先将 15 只染有肉瘤的小白鼠按体重大小配成 5 个区组,使每个区组内的 3 只小白鼠体重最接近,然后随机决定每个区组中的 3 只小白鼠接受 3 种药物的治疗,以肉瘤的重量为指标,试验结果见表 12-2。试比较不同抗癌药物对小白鼠肉瘤的抑瘤效果之间的差别是否有统计学意义?

(2)如何才能把同源配对设计拓展或镶嵌到单因素多水平设计中?

【例 12-25】 为了比较准确地评价分别含 5 种不同蛋白质饲料的营养价值,有人设计了如下的试验:用 8 窝大白鼠,每窝选取性别相同、体重接近的 5 只,用随机的方法将每窝中的 5 只分到 5 个饲料组中去,收集了试验数据见表 12-3。

表 12-2 不同药物作用后小白鼠肉瘤重量

区 组	肉瘤重量(g)		
	* :A	B	C
1	0.82	0.65	0.51
2	0.73	0.54	0.23
3	0.43	0.34	0.28
4	0.41	0.21	0.31
5	0.68	0.43	0.24

注:*代表“药物种类”

表 12-3 8 窝大白鼠分别用 5 种饲料喂养 9 周后所增体重

窝 别	体重增加量(g)				
	* :甲	乙	丙	丁	戊
1	24	15	37	57	82
2	42	28	37	51	66
⋮	⋮	⋮	⋮	⋮	⋮
7	47	21	32	54	73
8	53	37	42	59	90

注:*代表“饲料种类”

2. 具有一个重复测量的单因素设计

如何才能把自身配对设计拓展或镶嵌到单因素多水平设计中?

【例 12-26】 某外科医生在每一个患者的手术过程中的 3 个不同阶段测得患者门静脉压力资料见表 12-4,该资料所代表的试验属于具有一个重复测量的单因素设计。

表 12-4 10 名患者在手术过程中的 3 个不同阶段门静脉压力的测定结果

病 例 号	门静脉压力(cmH ₂ O)		
	* :A ₁	A ₂	A ₃
1	40	36	36
2	19	17	18
⋮	⋮	⋮	⋮
10	36	37	38

注:*代表“A(阶段)”

12.3.2 平衡不完全随机区组设计与双因素无重复试验设计

1. 平衡不完全随机区组设计

实际问题:在某试验研究中,若试验因素的水平数大于每个区组内可以容纳的受试对象个数时,应该如何处置呢?

表 12-5 四种方法治疗脚气的评分

受试者 编 号	疗 效 评 分			
	* :甲	乙	丙	丁
1	5	3
2	4	7
3	2	...	6	...
4	...	4	...	8
5	3	7
6	...	2	5	...

注:*代表“治疗方法”

【例 12-27】 用四种方法治疗脚气,受试者为两脚都患脚气的患者,每只脚接受一种处理,观察指标为治疗效果评分。设计格式和资料见表 12-5。

本例中, $v=4$ (药物有 4 种)、 $a=2$ (每位患者形成一个区组,每人只有两只脚), $r=3$ (每种药物被重复使用的次数), $b=6$ (区组个数,即本例中的患者人数),满足 $rv=ab$,即 $3 \times 4=2 \times 6$;同时满足 $\lambda=r(a-1)/(v-1)=3(2-1)/(4-1)=1$ 为整数。故本例为平衡不完全区组设计。受试对象是 6 名患脚气病的患者,试验因素是“治疗方法”,区组因素为每位患者(即个体差异),定量的观测指标为“疗效评分”。

2. 双因素无重复试验设计

实际问题：如何看待两个试验因素各水平组合条件下只做一次试验的结果？

【例 12-28】将落叶松苗木栽在四块不同的苗床(B)上，每块苗床的苗木又分别使用三种不同的肥料(A)以观察肥效差异，两年后于每一苗床的各施肥小区内用随机方法抽取一株，测其高度，资料见表 12-6。

特别提示：基于此设计所得出的结论是否正确取决于以下两点：其一，两试验因素之间的交互作用可忽略不计；其二，各试验条件下的试验结果十分稳定，其误差在专业上允许的范围之内。

表 12-6 不同肥料 A 和不同苗床 B 对落叶松苗高影响的测定结果

肥料 A	苗 高 (cm)				
	* :B ₁	B ₂	B ₃	B ₄	合计
A ₁	55	47	47	53	202
A ₂	63	54	57	58	232
A ₃	52	42	41	48	183
合计	170	143	145	159	617

注：* 代表“苗床 B”

12.3.3 拉丁方设计与交叉设计

1. 拉丁方设计

实际问题：如何才能使用最少的样本含量获得比较准确的试验结果？

【例 12-29】试评价 4 台秤测定结果是否一致，样品重量为观测指标，设计和资料见表 12-7，此为 4×4 拉丁方设计。

表 12-7 4 台秤称样品重量的一致性试验

样 品 代 号	秤的代号 (样品重量, mg)			
	* :1	2	3	4
1	C()	B()	A()	D()
2	B()	A()	D()	C()
3	A()	D()	C()	B()
4	D()	C()	B()	A()

注：* 代表“试验顺序”

特别提示：表中括号内填写样品重量；所选取的 4 个样品的重量最好不等，且包含实际使用时重量的最小值和最大值，这样有利于评价结果在一定重量范围内的准确性。

2. 2×2 交叉设计

(1) 成组 2×2 交叉设计。

【例 12-30】某麻醉科医生研究催醒宁对氟呱啉的作用，选用生理盐水(记为 C)作为催醒宁(记为 T)的对照药，拟用大鼠 6 只作为受试对象。先将 6 只大鼠随机均分成两组，每组 3 只；再用随机方法决定第一组 3 只大鼠接受两种处理的顺序，另一组中的 3 只大鼠接受两种处理的顺序相反。设计与资料见表 12-8，此为成组 2×2 交叉设计。

表 12-8 考察催醒宁对氟呱啉的作用的试验结果

大 鼠 编 号	是否用催醒宁与出现蹲和走动的时间(单位)	
	* :第 1 次	第 2 次
1	T(15)	C(27)
3	T(17)	C(28)
6	T(5)	C(26)
2	C(31)	T(25)
4	C(30)	T(14)
5	C(25)	T(18)

注：* 代表“试验顺序”

(2) 配对 2×2 交叉设计。

【例 12-31】 为研究 A(90402 中药复方)、B(安慰剂)两种处理对提高高原劳动能力的影 响,以条件近似的 10 名健康人作为受试对象,把受试对象和试验时间(冬季、春季)作为两个重要的非处理因素,每人都服用两种药物各一次。试验设计方案如下:将条件最接近的每两人配成一对,用随机的方法确定每对中之使用 A、B 药物的顺序,另一个人服药的顺序相反。设计格式和试验结果列在表 12-9 中,其观测指标为 PWC170(即把心跳校正到 170 次/min 时能做的功),此为配对 2 × 2 交叉设计。

表 12-9 两种药物对提高高原劳动能力的试验结果

受试者编号	药物与 PWC170		受试者编号	药物与 PWC170	
	春 季	* :冬 季		* :冬 季	春 季
1	A 159.4	B 153.8	6	B 129.9	A 146.1
2	B 129.4	A 159.8	7	B 166.3	A 210.9
3	B 122.1	A 137.6	8	A 208.7	B 169.9
4	A 130.4	B 137.8	9	A 180.1	B 150.7
5	A 150.7	B 140.2	10	B 143.4	A 150.8

注: * 代表“试验时间”,(1, 2)、…(9, 10)号分别为配对组; A(90402 中药复方)、B(安慰剂)。

12.3.4 嵌套(系统分组)设计、裂区(分割)设计及重复测量设计

1. 嵌套(系统分组)设计

(1) 试验因素存在自然属性上的嵌套关系。

【例 12-32】 嵌套设计的第一个例子:为研究萝卜叶子中 M 物质的含量,今随机采集了三棵萝卜(设为 A₁、A₂、A₃),各棵萝卜中随机取几片叶子(设为 B₁、B₂、B₃)作测定,结果见表 12-10。显然, M 物质含量的多少,不仅取决于所取自的叶子,更主要地取决于所取自的萝卜,即不同萝卜之间的变异大于同一棵萝卜叶子之间的变异。

表 12-10 萝卜叶子中 M 物质的测定结果

编 号	M 物质含量(单位)						
	A ₁ (B ₁	B ₂)	A ₂ (B ₁	B ₂	B ₃)	A ₃ (B ₁	B ₂)
1	3.28	3.52	2.46	1.87	2.19	2.77	3.74
2	3.09	3.48	2.44	1.92	2.19	2.66	3.44
3	3.31	4.07	.	2.10	.	.	.

(2) 试验因素对定量结果指标的影响存在主次关系。

【例 12-33】 在某项化合物的试验研究中,涉及催化剂的种类(因素 A)和温度(因素 B),由专业知识可知,对该化合物转化率的影响,催化剂的作用大于温度,而且,各催化剂条件下所用的温度不完全相同,设计格式和资料见表 12-11。

表 12-11 不同催化剂在不同温度下对某化合物转化率影响的观测结果

试验批次	转化率(%)								
	* :甲(70 80 90)			乙(55 65 75)			丙(90 95 100)		
1	82	91	85	65	62	56	71	75	85
2	84	88	83	61	59	60	67	78	89

注: * 代表“催化剂与温度(℃)”

2. 裂区(分割)设计

(1) 试验因素施加于受试对象时存在先后顺序之分。

【例 12-34】 嵌套设计的第一个例子：为研究不同瘤株的生瘤效果和不同浓度蛇毒的抑瘤作用，先将 48 只小鼠按原始体重分成 3 个随机区组（注：每个区组内的 16 只小鼠体重最接近），将每个区组内的 16 只小鼠随机地均分成 4 组，分别接种 4 种不同的瘤株，观察肿瘤生长情况。1d 后再对接同一种瘤株的 4 只小鼠分别腹腔注射 4 种不同浓度的蛇毒，连续用蛇毒抑瘤 10d，停药 1d 后解剖测瘤重。设计格式和资料见表 12-12，此为裂区（或分割）设计。

注意：在本例中，瘤株先作用于受试对象、蛇毒后作用于受试对象。

（2）试验过程分批完成。

【例 12-35】 嵌套设计的第二个例子：在某项试验中，涉及药物种类和药物剂量，但试验需要在不同时间重复进行，显然，时间改变是较慢的，而在每个确定的时间点上，药物种类与剂量的水平改变都是很方便的。若用大鼠作为受试对象，可选取来自大鼠的重要非试验因素作为子区组因素，还可以不同天做试验作为大区组因素，将药物种类与剂量进行组合构成析因结构，在以天为大区组因素的各水平下，再以时间段和时刻为两层子区组因素，按含区组因素析因设计分配大鼠。设计格式和资料见表 12-13。

表 12-12 不同瘤株与不同浓度的蛇毒共同作用后对小鼠抑瘤效果的影响

瘤株 A	蛇毒浓度 B	瘤重(g)		
		* :1	2	3
A ₁ (S180)	B ₁	0.80	0.76	0.36
	B ₂	0.36	0.26	0.31
	B ₃	0.17	0.28	0.16
	B ₄	0.28	0.13	0.11
A ₂ (HS)	B ₁	0.74	0.43	0.57
	B ₂	0.50	0.46	0.32
	B ₃	0.42	0.20	0.20
	B ₄	0.36	0.26	0.32
A ₃ (EC)	B ₁	0.31	0.55	0.32
	B ₂	0.20	0.15	0.20
	B ₃	0.38	0.18	0.26
	B ₄	0.25	0.21	0.14
A ₄ (ARS)	B ₁	0.48	0.57	0.33
	B ₂	0.18	0.30	0.29
	B ₃	0.44	0.27	0.27
	B ₄	0.22	0.30	0.37

注：* 代表“随机区组”

表 12-13 不同药物、剂量和作用时间下大鼠胸膜腔内积液量

日期	时段	时刻	剂量	胸膜腔内积液量(\bar{x} , ml, n=5)							
				* :1	2	3	4	5	6	7	8
第一天	上午	1	1	5.7	8.6	6.9	6.6	6.7	7.4	5.7	6.7
		1	2	5.1	7.2	6.8	6.4	6.6	8.7	6.7	7.0
		2	1	8.4	9.6	9.3	11.1	121.5	8.7	9.3	9.5
		2	2	7.3	8.7	7.9	6.9	8.9	9.5	8.3	11.3
	:	:	:	:	:	:	:	:	:	:	:
第二天	下午	1	1	5.8	6.8	7.0	8.5	7.8	7.3	6.4	8.5
		1	2	5.4	7.9	8.0	6.4	8.4	7.1	6.4	7.2
		2	1	9.1	10.8	6.9	12.2	9.9	10.4	10.6	10.5
		2	2	5.3	10.4	8.2	8.1	10.9	9.8	8.4	14.6
	:	:	:	:	:	:	:	:	:	:	:

说明：* 代表“药物编号”

第一行“省略号”代表省略了第一天下午的四行结果，第二行“省略号”代表省略了第二天上午的四行结果。

提示：本例中，“时间因素”有三个不同名称，即“天”、“时段”和“时刻”，前者包含后者，它们都可被视为“区组因素”，而真正的试验因素“药物种类”与“剂量”。相当于把“药物种类”与“剂量”所决定的两因素析因设计分别在“时刻”、“时段”和“天”上进行重复。表 12-13 中的每个数据代表一个平均值，统计分析时应给出全部原始数据。

3. 重复测量设计

(1)“时间”为重复测量因素。

①具有一个重复测量的单因素设计。

【例 12-36】 某外科医生在每一个患者的手术过程中 10 名患者在手术过程中的 3 个不同阶段测得患者门静脉压力资料见表 12-14，该资料所代表的试验属于具有一个重复测量的单因素设计。

表 12-14 10 名患者在手术过程中的 3 个不同阶段门静脉压力的测定结果

病 例 号	门静脉压力 (cmH ₂ O)		
	* :A ₁	A ₂	A ₃
1	40	36	36
2	19	17	18
⋮	⋮	⋮	⋮
10	36	37	38

注：* 代表“A(阶段)”

②具有一个重复测量的三因素设计。

【例 12-37】 某研究者为了探讨两种不同处理(因素 A)和不同剂量(因素 B)作用下不同时间点(因素 T)上 NFS-60 细胞凋亡情况，将全部样品随机地均分为 6 个组，每组含 3 个样品。对各组中的每个样品在处理后的 5 个不同时间点上观测 200 个细胞，数出凋亡细胞个数。因素 A 分为 A₁(单纯照射)、A₂(照射 + 过钒酸钠(Perv))，各处理组的照射剂量(因素 B, Gy)又分为 0、3、10。设计格式和资料见表 12-15，该资料为具有一个重复测量的三因素设计。

表 12-15 不同处理和不同剂量作用下不同时间点上 NFS-60 细胞凋亡情况

处理组 (因素 A)	照射剂量 (Gy)	样品 编号	200 个细胞中凋亡细胞数				
			* :T ₁ (12)	T ₂ (24)	T ₃ (36)	T ₄ (48)	T ₅ (60)
A ₁ (照射)	0	1	3	7	7	10	17
		2	5	3	5	7	18
		3	5	6	8	14	14
	3	1	4	7	10	20	33
		2	3	4	14	22	37
		3	3	5	9	19	31
	10	1	6	18	38	83	101
		2	4	16	42	91	120
		3	5	18	45	92	117
A ₂ (加 Perv)	0	1	4	6	4	12	14
		2	2	4	6	14	17
		3	3	7	3	11	19
	3	1	5	7	7	14	17
		2	3	5	9	12	19
		3	7	6	6	15	16
	10	1	5	10	16	33	49
		2	4	12	12	37	44
		3	3	9	14	36	52

注：* 代表“时间 T(h)”

(2)“处理”与“部位”组合起来构成复合型重复测量因素。

【例 12-38】 用贲门癌患者的标本制成液体，在三种不同处理条件下观测鸡胚背根神经节与鸡胚交感神经节中长出突起的神经节的比例。现有贲门癌患者 10 例，将每人的标本均分成三份，分别给予三种不同的处理(因素 A)，即 A₁(加入 100ng/ml 神经生长因子)、A₂(加入 200ng/ml 神经生长因子)和 A₃(单用贲门癌培养液)；并对每种处理后的标本中的两种类型的神经节(因素 B)，

即 B_1 (背根神经节) 与 B_2 (交感神经节), 观测长出突起的神经节的比例 (Y)。设计格式和资料见表 12-16, 该资料为具有两个重复测量的两因素设计。

表 12-16 贲门癌患者的标本经三种处理后两种神经节中长出突起的神经节的比例

病 例 号	Y (长出突起的神经节的比例)					
	$A_1 (B_1)$	B_2	$A_2 (B_1)$	B_2	$A_3 (B_1)$	B_2
1	0.50	0.43	0.50	0.43	0.80	0.50
2	0.63	0.38	0.55	0.50	0.71	0.63
3	0.50	0.50	0.54	0.50	0.83	0.67
4	0.43	0.43	0.50	0.38	0.70	0.57
5	0.50	0.40	0.50	0.57	0.70	0.71
6	0.44	0.38	0.63	0.50	0.77	0.63
7	0.43	0.50	0.50	0.38	0.69	0.50
8	0.63	0.44	0.40	0.38	0.70	0.50
9	0.44	0.50	0.50	0.44	0.67	0.50
10	0.44	0.43	0.45	0.40	0.60	0.75

(3)“处理”与“时间”组合起来构成复合型重复测量因素。

【例 12-39】 某研究者为了研究阿托品累计用量对有机磷药物中毒后的治疗效果, 将 10 只大鼠随机均分成两组, 第一组 5 只接受有机磷农药使之中毒, 第二组 5 只未接受有机磷农药。两组大鼠均给予一次乙酰胆碱, 同时, 4 次给予阿托品治疗, 每次均为 2mg/kg, 在每次给予阿托品治疗后的 2min、5min、10min 观测膈肌对电刺激产生的反应, 用“面积”表示, 面积越大, 治疗效果越好。设计和资料见表 12-17, 该资料为具有两个重复测量的三因素设计。

表 12-17 连续 4 次给阿托品治疗不同时间点上膈肌对电刺激产生反应的观测结果

中毒与否(A)	大鼠编号	面 积					
		B 与 C: B_1 (2 5 10)			B_2 (2 5 10)		
是	1	48.0	48.6	51.0	52.8	50.6	49.7
	2	45.1	47.7	48.7	47.1	45.7	42.3
	3	45.1	47.7	48.7	46.0	46.0	43.7
	4	41.4	41.1	42.0	39.1	38.0	35.4
	5	57.7	60.6	61.4	58.6	56.8	52.0
否	1	64.8	66.3	67.1	66.0	65.1	62.0
	2	59.7	59.7	59.1	59.4	59.1	59.4
	3	53.1	54.6	54.6	52.3	51.7	49.1
	4	61.1	62.9	60.6	56.8	52.8	47.1
	5	84.6	85.1	84.3	80.3	77.1	72.3

注: 表中右边还有 6 列数据未列出, 其表头为“ B_3 (2 5 10) min”和“ B_4 (2 5 10) min”。

12.3.5 析因设计及含区组因素的析因设计

1. 析因设计

(1)何为析因设计。具备下列 7 个特点的多因素试验设计可被称为析因设计:

第一, 试验因素的个数 ≥ 2 ;

第二, 每个试验因素的水平数 ≥ 2 且可以不等;

第三, 不同的试验条件数(或组合数或试验点数)等于全部试验因素的水平数之乘积;

第四，各试验条件下至少要做两次独立重复试验；

第五，全部受试对象被完全随机地分配进入任何一个试验条件组中，各小组中的受试对象个数可以不等，但最好相等；

第六，试验时，任何一次试验都将涉及任何一个试验因素的某个水平，即全部试验因素同时施加；

第七，进行统计分析时，假定全部试验因素对观测结果的影响是地位平等的，即没有主次之分。

由上述特点，决定了析因设计分别具有一个明显的优点和缺点。优点：可以分析每个试验因素的主效应和试验因素之间各级交互作用的效应大小。缺点：在试验因素个数多和/或水平数多时，总试验次数太多，试验费用增大且费时费力。

(2)析因设计的实例。

【例 12-40】 给 40 只雌性小鼠完全随机地均分为 4 组，每组 10 只。第一组为空白对照组，即不给予任何药物治疗；第二组为仅给予一定剂量的激素治疗；第三组为仅给予一定剂量的某种中药治疗；第四组为既给予激素又给予中药(剂量同前面两组)治疗。观测指标为“排卵数”。设计格式见表 12-18，此为两因素析因设计(具体为 2×2 析因设计)。

【例 12-41】 某试验同时涉及 A、B、C 3 个地位平等的试验因素，A 分为 2 个水平、B 分为 3 个水平、C 分为 4 个水平，观测指标为 OD 值，受试对象为样品，不同试验条件下均独立地重复做了 2 个样品，资料见表 12-19，此为三因素析因设计(具体为 $2 \times 3 \times 4$ 析因设计)。

表 12-18 中药用否与激素用否对雌性小鼠排卵数的影响情况

中 药 用 否	排卵数(个, $\bar{x} \pm s$)	
	* :用	否
用	38.40 \pm 26.91	10.10 \pm 5.32
否	37.90 \pm 27.18	8.90 \pm 5.20

注：* 代表“激素用否”

表 12-19 3 个试验因素作用下 OD 值的测定结果

因素 A 与 B		OD 值							
		因素 C: C ₁		C ₂		C ₃		C ₄	
A ₁	B ₁	0.39	0.41	0.37	0.39	0.42	0.38	0.44	0.41
	B ₂	0.37	0.36	0.43	0.45	0.41	0.37	0.42	0.39
	B ₃	0.45	0.43	0.46	0.39	0.38	0.35	0.39	0.37
A ₂	B ₁	0.36	0.41	0.45	0.36	0.41	0.45	0.41	0.46
	B ₂	0.42	0.37	0.38	0.41	0.38	0.36	0.43	0.38
	B ₃	0.37	0.43	0.36	0.39	0.43	0.42	0.35	0.37

2. 含区组因素的析因设计

(1)何为含区组因素的析因设计。

先将受试对象按重要非试验因素形成区组(每个区组中受试对象的个数等于试验因素的水平组合数)，再将每个区组中的受试对象完全随机地均分入全部试验因素水平组合中去。

(2)含区组因素的析因设计的实例。

【例 12-42】 有人研究猪食用不同饲料对体重增加量的影响。A(大豆粉 + 不同含量的蛋白质)：A₁(加 14% 蛋白质)、A₂(加 12% 蛋白质)；B(玉米中乙氨酸的含量)：B₁(含 0.6% 乙氨酸)、B₂(缺乏乙氨酸)，共有 4 种不同的饲料配方。用 24 头猪作为受试对象，按猪的初始体重由轻到重

排序,并形成 6 个随机区组,每组中的 4 只猪体重接近,被随机地分入 4 个饲料组中去,在喂养期间,尽可能使每只猪每餐都能吃饱。设计格式和资料见表 12-20,该资料所取自的试验设计类型为含区组因素的析因设计。

表 12-20 两种饲料的不同配方对猪平均日增重量的影响结果

随机区组 编号	平均日增重量(×500g)			
	A ₁ (B ₁)	B ₂)	A ₂ (B ₁)	B ₂)
1	1.38	1.52	1.22	1.11
⋮	⋮	⋮	⋮	⋮
6	1.47	1.53	1.16	0.99

12.3.6 多因素试验设计类型的辨析

1. 辨析多因素设计类型的意义

对于一个全面掌握多因素试验设计类型的科研人员来说,每次进行试验之前都制订出了科学完善的试验设计方案,具体的设计类型已明确写在方案中,并按此方案具体实施。到处理数据时,不存在判断设计类型的问题。然而,当数据处理者本人并不十分清楚所面对的数据究竟属于什么设计类型,为了合理选用统计分析方法处理给定的资料,正确判定定量资料所对应的试验设计类型就是一件十分重要的事。可以说,不判定或错误判定了拟处理的定量资料所对应的设计类型,误用统计分析方法得出错误结论将是不可避免的。

2. 辨析多因素设计类型的技巧

面对一个较复杂的实际资料,如何辨析其所对应的多因素设计类型呢?当反映试验因素及其水平和试验结果的表格中,已明确写出“正交设计”或“均匀”等字样时,设计类型已明确告知,不存在辨析的问题;若仅写出了一些试验条件和各试验条件下的试验结果时,就必须先辨析其试验设计类型。要想使判断结果正确,最重要的是要弄清此试验是怎样做的;其次,应提炼出全部试验因素(用专业名称,如药物种类、剂量大小、作用时间等;不要用“组别”、“处理”和“方案”这些涵盖面很宽的“集合词”)及其水平;再次,应弄清试验因素对观测结果的影响有无主次之分、设计中有无区组因素(即重要的非试验因素);最后,当试验中涉及“时间”或“部位”因素时,务必要仔细考察具体试验是怎样做的,弄清它们是否为“重复测量因素”。若在它们的不同水平下,对各组的每个受试对象或样品在同一个定量指标上进行多次观测,得到取值,那么,它们就被称为“重复测量因素”,所对应的设计应被称为“重复测量设计”。

有些从表面上看是“多因素非平衡组合试验资料”,但仔细斟酌,对其列表格式重新布局,发现它却是一个标准的多因素设计类型。对这样的资料,正确判定其设计类型显得尤为重要。

3. 改变列表格式突显设计类型

【例 12-43】 某研究者欲比较经心外膜、静脉移植骨髓间充质干细胞(MSCs)治疗大鼠急性心肌梗死(AMI)的效果,为临床选择移植方法提供依据。选择 SD 大鼠,建立 AMI 模型,取造模成功的大鼠 60 只,随机等分为 4 组,24h 内进行移植。其中,静脉对照组大鼠尾静脉注射 L-DMEM 0.7ml;静脉移植组大鼠尾静脉注射 MSCs 悬液 0.7ml;心外膜对照组大鼠心肌梗死区中央及周边注射 L-DMEM 0.7ml;心外膜移植组大鼠心肌梗死区中央及周边注射 MSCs 悬液 0.7ml。观察大鼠呼吸、心跳平稳后关胸,术后正常喂养。3 周后用 VIVID-7 彩色多普勒超声诊断仪检测,然后麻醉进行血流动力学检测,记录左心室收缩压(LVSP)、左心室舒张末期压(LVEDP)、左室射血分数(LVEF)、短轴缩短率(FS)等,结果见表 12-21。组间比较采用 *t* 检验。

【对错误的辨析与释疑】 在表 12-21 中,“组别”之下有 4 个组,表面上看,该资料很像单因素四水平设计定量资料,但实则不然。表 12-21 只是实际工作者表达试验方法和定量资料的“表现型”。而这个表现型与它的原型和标准型不一致,人们很容易被假象所迷惑,直接去套用统计

学教科书上的标准型，出错就成为必然的事了。事实上，仔细观察 4 种不同的实验条件，可发现它包含两个实验因素：注射部位和药物种类，两者均为两水平，前者水平为是或否，后者水平为静脉或心外膜。4 种不同的实验条件恰为这两个因素水平的全面组合，与此问题对应的“原型（同时也是本问题的标准型）”如表 12-22 所示。由于研究者在每种实验条件下均进行了 15 次独立重复实验，因此若无专业依据、预试验或文献资料表明两因素对观测指标影响的重要程度存在差异时，本资料应为两因素析因设计定量资料。分析时，应采用两因素析因设计定量资料的方差分析来处理。

表 12-21 4 组大鼠超声结果及血流动力学比较 ($\bar{x} \pm s$) (表现型)

组 别	LVEF (%)	FS (%)	LVSP (kPa)	LVEDP (kPa)
0 静脉对照组	22.14 ± 3.23	18.23 ± 2.15	12.00 ± 0.54	2.85 ± 0.28
心外膜对照组	22.59 ± 2.85	18.18 ± 2.32	12.05 ± 0.66	2.88 ± 0.29
静脉移植组	32.21 ± 2.68	23.32 ± 3.78	13.46 ± 0.78	2.30 ± 0.22
心外膜移植组	40.11 ± 3.27	27.56 ± 4.36	15.41 ± 0.88	1.89 ± 0.56

表 12-22 4 组大鼠超声结果及血流动力学比较 ($\bar{x} \pm s$) (原型和标准型)

注 射 部 位	药 物 种 类	LVEF (%)	FS (%)	LVSP (kPa)	LVEDP (kPa)
静 脉	L-DMEM	22.14 ± 3.23	18.23 ± 2.15	12.00 ± 0.54	2.85 ± 0.28
	MSCs	32.21 ± 2.68	23.32 ± 3.78	13.46 ± 0.78	2.30 ± 0.22
心外膜	L-DMEM	22.59 ± 2.85	18.18 ± 2.32	12.05 ± 0.66	2.88 ± 0.29
	MSCs	40.11 ± 3.27	27.56 ± 4.36	15.41 ± 0.88	1.89 ± 0.56

【例 12-44】 某人将 40 只小鼠随机地均分成 4 个组，用 6.5Gy 某种射线照射小鼠后，分别给予不同的处理，即第一组，空白对照；第二组，用 rhIL-6；第三组，用 rhG-CSF；第四组，用 rhIL-6 + rhG-CSF。每组小鼠均在照射前、照射后 7d、10d、14d、17d、21d、23d 7 个不同时间点上被重复观测白细胞的数值，设计格式和资料见表 12-23。原作者用多次 t 检验处理此资料，这是错误的！欲选用正确的统计分析方法，首先需正确判断该资料所取自的设计类型。

表 12-23 6.5Gy 某种射线照射小鼠外周血白细胞数的动态变化情况 ($\bar{x} \pm s$) (表现型)

照后 天数	WBC ($\times 10^9/L$)			
	对照	rhIL-6	rhG-CSF	rhIL-6 + rhG-CSF
照前	19.7 ± 1.8	17.3 ± 3.2	17.9 ± 3.8	18.3 ± 3.7
7	1.8 ± 0.7	1.9 ± 0.8	2.0 ± 1.0	2.2 ± 1.5
10	1.8 ± 0.9	1.7 ± 0.8	1.7 ± 0.6	1.3 ± 0.6
14	1.9 ± 0.6	2.9 ± 1.0	2.7 ± 1.1	2.5 ± 0.9
17	2.8 ± 1.0	3.3 ± 0.9	4.6 ± 1.2	4.5 ± 1.1
21	7.6 ± 2.0	6.6 ± 2.4	6.0 ± 3.0	6.9 ± 4.2
23	9.4 ± 3.8	9.7 ± 4.2	9.7 ± 3.9	8.7 ± 5.6

注:rhIL-6, 200μKD; rhG-CSF, 125μKD; $n=10$ 。

【对差错的辨析与释疑】 在表 12-23 中，究竟涉及了几个试验因素？很多人都认为它涉及了两个因素，一个是药物种类，另一个是照射后时间。其实，它涉及了三个试验因素，除照射后时间外，纵向所列的四组是各有 2 水平的两个因素的 4 种组合，并非是一个药物因素的四个水平。仔细观察不难看出，这个组合因素的四个水平就是由“rhIL-6 不用与用”、“rhG-CSF 不用与用”的四种组合。将这四种组合视为四个试验条件，每个试验条件下的 10 只小鼠在 7 个不同的时间点上被重

复观测 WBC 的值, 故与表 12-23 对应的试验设计类型应叫做“具有一个重复测量的三因素设计”。若将该表资料用表 12-24 的形式表达, 则容易辨别出它所对应的设计类型。

表 12-24 表 12-23 转换形式后的结果 ($\bar{x} \pm s$) (原型和标准型)

rhIL-6 使用与否	rhG-CSF 使用与否	WBC ($\times 10^9/L$)						
		* : 照前	7	10	14	17	21	23
a ₁ (不用)	b ₁ (不用)	19.7 \pm 1.8	1.8 \pm 0.7	1.8 \pm 0.9	1.9 \pm 0.6	2.8 \pm 1.0	7.6 \pm 2.0	9.4 \pm 3.8
	b ₂ (用)	17.3 \pm 3.2	1.9 \pm 0.8	1.7 \pm 0.8	2.9 \pm 1.0	3.3 \pm 0.9	6.6 \pm 2.4	9.7 \pm 4.2
a ₂ (用)	b ₁ (不用)	17.9 \pm 3.8	2.0 \pm 1.0	1.7 \pm 0.6	2.7 \pm 1.1	4.6 \pm 1.2	6.0 \pm 3.0	9.7 \pm 3.9
	b ₂ (用)	18.3 \pm 3.7	2.2 \pm 1.5	1.3 \pm 0.6	2.5 \pm 0.9	4.5 \pm 1.1	6.9 \pm 4.2	8.7 \pm 5.6

注: * 代表“时间(d)”

4. 多因素非平衡组合试验的拆分与重组

【例 12-45】 某研究者欲探讨粒细胞集落刺激因子(G-CSF)对大鼠冠状动脉微栓塞(CME)后心肌细胞凋亡的影响以及 Janus 激酶-信号转导子与转录激活子(JAK/STAT)通路的介导作用, 选择 92 只雄性成年 SD 大鼠, 随机分为 CME 组(24 只)、G-CSF 组(24 只)、JAK2 特异性抑制剂 AG490 组(G-CSF + AG490, 24 只)和假手术组(20 只)。CME 组、G-CSF 组及 AG490 组升主动脉夹闭后自左室腔内注入自体微血栓, 造成 CME, 假手术组注入等量生理盐水。G-CSF 组和 AG490 组术后 2h 起给予皮下注射重组人 G-CSF(rhG-CSF) 100 $\mu\text{g} \cdot \text{kg}^{-1} \cdot \text{d}^{-1}$ 持续 5d, AG490 组同时给予 AG490 溶液腹腔注射 5 $\mu\text{g} \cdot \text{kg}^{-1} \cdot \text{d}^{-1}$, 其他组给予等量生理盐水。术后 3d、1、2 及 4 周处死动物。各组心肌样品中以实时定量聚合酶链式反应法检测 Bcl-2、Bax、Fas 及 FasL 的 mRNA 表达, 并计算 Bcl-2/Bax 比值; 脱氧核糖核苷酸末端转移酶介质的缺口末端标记(TUNEL)法测定凋亡细胞, 结果见表 12-25。统计分析采用 *t* 检验或方差分析。

表 12-25 各组大鼠心肌细胞凋亡指数的比较 ($\bar{x} \pm s$)

组 别	鼠数(只)	3d	1 周	2 周	4 周
CME 组	6	11.5 \pm 1.7	17.2 \pm 1.9	8.3 \pm 1.4	4.6 \pm 0.8
G-CSF	6	4.5 \pm 0.5	6.1 \pm 1.0	4.1 \pm 0.6	2.9 \pm 0.5
AG490	6	6.6 \pm 0.7	14.2 \pm 1.5	6.3 \pm 0.8	4.1 \pm 0.7
假手术组	6	1.2 \pm 0.4	1.2 \pm 0.6	1.5 \pm 0.7	1.4 \pm 0.6

【对差错的辨析与释疑】 研究者将 92 只雄性成年 SD 大鼠随机分为四组: 假手术组、CME 组、G-CSF 组和 AG490 组。其中, CME 组、G-CSF 组和 AG490 组均建立 CME 模型, 假手术组未建立 CME 模型, 故是否建立 CME 模型是一个实验因素。相对于 CME 组, G-CSF 组和 AG490 组又给予了额外的处理, 即皮下注射重组人 G-CSF 5d, 所以是否注射 G-CSF 也是一个实验因素。AG490 组在注射 G-CSF 的基础上同时给予 AG490 溶液腹腔注射, 故是否注射 AG490 溶液也是一个实验因素。此外, 研究者在术后 3d、1、2 及 4 周分别处死一定数量的动物, 所以处死时间也是一个实验因素。而每只动物只是在某一个时间点上进行了观测, 故本资料中时间因素不是重复测量因素。

关于时间因素的性质, 是一个需要注意的问题。若研究者在多个时间点上对同一只受试对象进行重复观测, 从而获得同一指标的多个观测值, 则时间因素属于重复测量因素。反之, 若研究者在多个时间点上观测不同的受试对象, 从每个受试对象身上只可能获得某指标的一个观测值, 也就是说每个受试对象只能在若干个时间点中的一个时点上被观测, 则时间为一般的实验因素, 不属于重复测量因素。

综上所述, 本资料共涉及 4 个实验因素: 是否建立 CME 模型、是否注射 G-CSF、是否注射

AG490 溶液和处死时间,各自水平数分别为 2、2、2 和 4。这四个因素水平的全面组合应有 32 种 ($2 \times 2 \times 2 \times 4 = 32$) 实验条件,但实际上研究者只在其中的 16 种条件下进行了实验,所以本资料应为多因素非平衡组合实验定量资料。分析时,应根据专业知识和统计学知识,对资料进行合理的拆分组合,然后根据拆分后的各种组合所对应的设计类型选择相应的统计分析方法。对本资料而言,可拆分组合如下。

组合 1: CME 组和假手术组在 4 个时间点上的数据。这个组合涉及两个实验因素:是否建立 CME 模型和处死时间,且此组合的实验条件为两因素各水平全面组合而成,各实验条件下均进行了多次独立重复实验,故组合 1 应为两因素析因设计定量资料。

组合 2: CME 组和 G-CSF 组在 4 个时间点上的数据。这个组合涉及两个实验因素:是否注射 G-CSF 和处死时间,且此组合的实验条件为两因素各水平全面组合而成,各实验条件下均进行了多次独立重复实验,故组合 2 应为两因素析因设计定量资料。

组合 3: G-CSF 组和 AG490 组在 4 个时间点上的数据。这个组合涉及两个实验因素:是否注射 AG490 溶液和处死时间,且此组合的实验条件为两因素各水平全面组合而成,各实验条件下均进行了多次独立重复实验,故组合 3 应为两因素析因设计定量资料。

可见,3 种组合资料的类型均为两因素析因设计定量资料,在满足参数检验的前提条件下,应采用两因素析因设计定量资料方差分析处理这些组合资料。研究者没有认真分析资料类型,随意套用 t 检验和方差分析的做法是不可取的。且研究者关于 t 检验和方差分析的说法太过笼统, t 检验包括单组设计定量资料 t 检验、配对设计定量资料 t 检验和成组设计定量资料 t 检验,方差分析包括单因素多水平设计定量资料方差分析、随机区组设计定量资料方差分析、析因设计定量资料方差分析等十多种,具体是何种 t 检验和方差分析,研究者并没有说明。这种现象在论文中较为常见,望引起科研人员的注意。

【例 12-46】 某研究者欲探讨烟酸对高脂血症兔脂肪细胞胆固醇流出和肝 X 受体(LXR) α 、过氧化物酶体增殖物激活受体(PPAR) γ mRNA 表达的影响,选取 12 只健康雄性新西兰兔给予高胆固醇饲料饲养 8 周,建立高胆固醇模型后,随机分为高胆固醇组(继续饲以高胆固醇饲料 6 周, $n = 6$),烟酸治疗组(在饲以高胆固醇饲料的基础上给予烟酸缓释剂 $0.2 \text{ g} \cdot \text{kg}^{-1} \cdot \text{d}^{-1}$,共 6 周, $n = 6$)。另选普通饮食 14 周雄性新西兰兔($n = 6$)作为对照组。分别于饲养前、饲养至第 8 周和第 14 周末,称取动物体重,检测血脂,结果见表 12-26 与表 12-27。数据以 $\bar{x} \pm s$ 表示,并采用单因素多水平设计定量资料方差分析进行多组间比较,自身前后及两组间比较均采用 t 检验。

表 12-26 3 组兔第 0、8、14 周血脂的变化 ($\bar{x} \pm s$)

组 别	TC (mmol/L)			LDL-C (mmol/L)		
	0 周	8 周	14 周	0 周	8 周	14 周
对照组 ($n = 6$)	3.04 ± 0.41	3.32 ± 0.41	3.16 ± 0.61	1.24 ± 0.52	1.18 ± 0.44	1.12 ± 0.36
高脂血症组 ($n = 6$)	3.56 ± 0.83	23.44 ± 1.58	26.45 ± 2.15	1.72 ± 0.75	20.58 ± 1.79	23.38 ± 2.15
烟酸治疗组 ($n = 6$)	3.49 ± 0.53	24.52 ± 2.39	14.84 ± 2.16	1.54 ± 0.51	21.89 ± 2.41	12.38 ± 2.10

表 12-27 3 组兔第 0、8、14 周血脂的变化 (续) ($\bar{x} \pm s$)

组 别	TC (mmol/L)			LDL-C (mmol/L)		
	0 周	8 周	14 周	0 周	8 周	14 周
对照组 ($n = 6$)	0.74 ± 0.20	0.89 ± 0.07	0.93 ± 0.07	1.35 ± 0.08	1.40 ± 0.15	1.33 ± 0.26
高脂血症组 ($n = 6$)	0.83 ± 0.24	1.76 ± 1.20	1.98 ± 0.18	1.41 ± 0.21	1.58 ± 0.29	1.02 ± 0.08
烟酸治疗组 ($n = 6$)	0.78 ± 0.29	1.52 ± 0.37	1.47 ± 0.28	1.50 ± 0.24	1.67 ± 0.20	1.74 ± 0.23

【对差错的辨析与释疑】 研究者将 18 只新西兰家兔分为 3 组, 其中 6 只为对照组, 给予普通饮食, 另 12 只给予高胆固醇饲料饲养以建立高胆固醇模型。可见, 饲料种类是一个实验因素。12 只高胆固醇家兔又分为 2 组, 其中 6 只继续喂以高胆固醇饲料, 另 6 只在喂以高胆固醇饲料的基础上给予烟酸缓释剂, 所以两组在实验条件上存在着是否给予烟酸缓释剂的差别, 因此, 是否给予烟酸缓释剂也是一个实验因素。研究者在饲养前、饲养后第 8 周和第 14 周末, 均检测家兔的血脂指标, 故观测时间是一个重复测量因素。所以, 本资料共涉及三个实验因素, 其水平的全面组合应为 12 种($2 \times 2 \times 3$) 实验条件, 而研究者只在其中的 9 种条件下进行了实验, 故本资料应为多因素非平衡组合实验定量资料。

研究者的做法是先将每个时间点上的数据单独分析, 将其视单因素三水平设计定量资料, 采用单因素三水平设计定量资料方差分析处理此资料; 然后将每组受试对象在某两个时间点上被观测的数据视为自身前后配对设计定量资料, 采用配对设计定量资料 t 检验来分析。这样做是错误的, 它是将资料进行随意的切割, 并以最简单的统计分析方法来处理。

根据前述分析, 本资料为多因素非平衡组合实验定量资料, 合适的分析方法是根据专业知识和统计学知识对其进行拆分组合。当然, 这里所说的合理的拆分组合不是像研究者那样把资料切割成最简单的设计类型, 然后再进行分析。一般而言, 应尽量切割成完整性较好的复杂设计类型, 如重复测量设计、析因设计等, 这样可以更好地考察因素之间的交互作用, 并降低统计结论犯假阳性错误的概率。

本资料可拆分如下:

组合 1: 对照组和高脂血症组在三个时间点上的数据。此组合涉及两个实验因素: 饲料种类和观测时间, 其中后者为重复测量因素, 故组合 1 应为具有一个重复测量的两因素设计。

组合 2: 高脂血症组和烟酸治疗组在三个时间点上的数据。此组合涉及两个实验因素: 是否给予烟酸缓释剂和观测时间, 其中后者为重复测量因素, 故组合 2 也为具有一个重复测量的两因素设计。

对组合 1 和组合 2, 可采用具有一个重复测量的两因素设计定量资料的方差分析来处理。最好是以家兔饲养前(即观测时间为 0 周时)的指标值为“基础值”, 采用具有一个重复测量的两因素设计定量资料的协方差分析来处理。因所观测的指标均反应受试对象的血脂情况, 可将四个指标一起分析, 采用具有一个重复测量的两因素设计定量资料的四元协方差分析, 这样将得到更具概括性和综合性的结论。

12.4 如何用 SAS 产生多因素试验设计类型

试验设计是科研工作中极其重要的一步。一个好的试验设计方案能够使科研工作者事半功倍地完成既定的科研任务; 相反, 如果试验前制订的试验设计方案不合理甚至根本就没有制定试验设计方案而是随心所欲地去做科研, 那么往往是事倍功半且容易得出错误的结论。本节主要介绍如何借助 SAS 软件的相关过程来实现试验设计方案中的一个关键环节——多因素设计类型的确定与呈现。

需要说明的是, 本节程序基于 SAS 9.2 简体中文版编写, 如读者所用版本为英文版, 请将程序中相关中文修改为英文字符, 方可正确显示设计方案。

12.4.1 试验设计方案包含哪些内容

试验设计方案, 从广义上来讲, 就是在试验开始前, 根据专业知识和统计学知识, 结合研究者的研究目的, 制定出指导如何进行科研工作的计划, 它不仅包含专业设计(充分利用基本常识和专

业知识)，而且包含试验设计(正确运用三要素、四原则和设计类型)，甚至包含试验过程中的质量控制；从狭义上讲，就是在试验开始前，制定的一个包含全部因素各水平有机组合的试验因素与水平的架构，简称多因素试验设计类型。它清楚地表达出每次试验涉及的因素及其水平，各因素水平组合下是否需进行独立重复试验以及全部试验共需进行的次数。它相当于试验研究者在试验过程中如何进行每次具体试验的行动指南，有助于参与试验的全部研究者在试验过程中有章可依、有据可查，从而科学、有效地解决试验研究中的复杂问题。简而言之，广义的试验方案是如何成功地实现研究目标的一揽子计划；而狭义的试验方案仅仅是因素与水平的组合方式或架构。本节介绍的内容特指后者。

为说明制订好的试验设计方案的优点，读者可先略览一下【例 12-47】，下面给出【例 12-47】资料的试验设计方案，见表 12-28。

表 12-28 【例 12-47】资料试验设计方案

窝 别	小鼠编号[体重增加量(g)]		
	* : 营养素 1	营养素 2	营养素 3
第 1 窝	3()	1()	2()
第 2 窝	6()	5()	4()
第 3 窝	8()	9()	7()
第 4 窝	11()	10()	12()
第 5 窝	14()	15()	13()
第 6 窝	18()	17()	16()

注：* 代表“营养素种类”

试验设计方案的解释参见【例 12-47】，表格内各家兔编号后的“()”内，可填写试验的结果。可见，该表格简单易用，较为清晰地呈现了资料的设计类型，并确保试验结果记录有序。本节将介绍如何借助 SAS 软件来实现各种常用多因素设计类型。当然，为保证程序简单易懂，SAS 输出的设计类型与表 12-28 给出的形式会略有差别，但不影响设计类型的展现和试验结果的记录。本节将介绍的这些多因素试验设计类型基本上可称为“固定模式的多因素设计类型”；还有一些则可称为“灵活模式的多因素设计类型”，如正交设计、均匀设计、分式析因设计、反应曲面设计、混料设计等。前者一般不需要通过计算(仅通过目测或手工排列)就可获得，而后者需要通过较复杂的计算才可构造出来。

12.4.2 用 SAS 实现随机区组设计

【例 12-47】 欲研究甲、乙、丙 3 种营养素对小白鼠体重增加的影响，取 6 窝小白鼠，每窝 3 只，随机安排喂养甲、乙、丙 3 种营养素中的 1 种。8 周后，观察小白鼠体重的增加情况。请给出具体的设计方案。

【分析】 对小白鼠进行编号，第 1 窝编号范围为 1~3，第 2 窝编号范围为 4~6，第 3 窝编号范围为 7~9，第 4 窝编号范围为 10~12，第 5 窝编号范围为 13~15，第 6 窝编号范围为 16~18。SAS 程序如下，程序名为 jcyx12_47.sas。

```
% let seed =10000;          /* 1 */
% let block =6;              /* 2 */
% let treat =3;              /* 3 */
% let b_values = '第 1 窝' '第 2 窝'
'第 3 窝' '第 4 窝' '第 5 窝'
'第 6 窝';                  /* 4 */

data b;                      /* 9 */
  set a;
  subject = _n_;
run;
ods html style =analysis; /* 10 */
proc tabulate style = [width =100]
```



```

% let t_values = '营养素 1' '营养素 2'
'营养素 3'; /* 5 */
% let subject_name = '小鼠编号'; /* 6 */
% let block_name = '窝别'; /* 7 */
proc plan seed = &seed; /* 8 */
  factors block = &block ordered
  treat = &treat;
  output out = a
  block cvals = (&b_values)
  treat cvals = (&t_values);
run;

format = 2.0;
class block treat;
var subject;
table block = ' ', subject =
&subject_name*
  treat = ' ' * sum = ' ' / box = [label
= &block_name];
run;
ods html close;

```

【程序说明】 第 1 步至第 7 步为设置宏变量，依次为随机种子数、区组因素水平数、处理因素水平数、各区组名称、各处理名称、受试对象名称(后加编号二字)、区组因素名称，使用时仅需根据实际情况修改这 7 个宏变量的取值即可。第 4 步至第 7 步中，宏变量为字符型变量，其取值需以英文单引号括起来，尤其是第 4 步和第 5 步，同时赋以因素的多个水平值，此时每个水平都需以英文单引号括起来，各水平之间需留有一个或多个空格。第 8 步是调用 PLAN 过程将数字进行随机化排列，结果输出至数据集 sheji 中。第 9 步对数据集 sheji 进行操作，产生受试对象编号。第 10 步调用 tabulate 过程将设计方案以报表形式展现出来。

【主要输出结果及解释】

窝别	小鼠编号		
	营养素 1	营养素 2	营养素 3
第 1 窝	3	1	2
第 2 窝	6	5	4
第 3 窝	8	9	7
第 4 窝	11	10	12
第 5 窝	14	15	13
第 6 窝	18	17	16

第 1 窝的 3 只小鼠编号为 1、2、3，按顺序对应分别接受营养素 2、营养素 3、营养素 1 的喂养，其他可以此类推。以上设计方案中，每个小鼠编号后面均有较大空隙，可用来记录试验的结果。

注意：SAS 输出的网页格式结果中，单元格中的数字为右对齐，但复制至 Word 文档中，将自动转换为左对齐。

12.4.3 用 SAS 实现拉丁方设计

【例 12-48】 欲研究 4 种不同剂量的某药物注射后对血糖升高值的影响，研究者选择 4 个受试对象，每个对象静脉给予不同剂量的该药物各一次，研究者拟采用拉丁方设计进行此试验，请给出具体的设计方案。

特别说明：此试验在实际使用时应慎用，因为每个个体身上先后使用了 4 种不同剂量的某药物，而且，剂量的顺序被随机化了，在多个个体身上观察到每种剂量的疗效是随机的、非线性的，但对试验结果采取此种设计下定量资料方差分析处理时，假定其疗效是线性的，故其分析结果很可能是错误的。此处仅仅展示如何用 SAS 呈现此种试验设计类型。若不是几种药物或几种剂量，而是三台测定体重的仪器，且观测指标是“体重”，则这样设计的结果就非常准确了。以后，遇到类似情况，也应同样理解和对待，不再赘述。

【分析】 对 4 名受试对象进行编号，范围为受试对象 1 ~ 受试对象 4，4 个试验阶段分别设为

阶段1~阶段4, 药物剂量设为1~4(试验因素的水平必须为数字, 这是由后续程序中调用的 tabulate过程要求的)。SAS 程序如下, 程序名为 jcyx12_48. sas。

```
% let seed=10000;          /* 1 */
% let block=4;              /* 2 */
% let period=4;             /* 3 */
% let treat=4;              /* 4 */
% let b_values='受试对象 1' '受试对象 2'
'受试对象 3' '受试对象 4'; /* 5 */
% let p_values='阶段 1' '阶段 2'
'阶段 3' '阶段 4';        /* 6 */
% let t_values=1 2 3 4;    /* 7 */
% let subject_name='受试对象编号';
/* 8 */
% let treat_name='剂量';    /* 9 */
proc plan seed=&seed;       /* 10 */
  factors block=&block ordered period=&period ordered;
  treatments treat=&treat cyclic;
  output out=a
    block cvals=(&b_values) random
    period cvals=(&p_values) random
    treat nvals=(&t_values) random;
run;

ods html style=analysis; /* 11 */
proc tabulate style=[width=100]
format=2.0;
  class block period;
  var treat;
  table block=' ', treat=&treat_name*period
    =' '*sum=' '/box=[label=&subject_name];
run;
ods html close;
```

【程序说明】 第1步至第9步为设置宏变量, 依次为随机种子数、区组因素水平数、试验阶段因素水平数、处理因素水平数、各区组名称、各阶段名称、各处理名称、受试对象名称(后加编号二字)、处理因素名称, 使用时只需根据实际情况修改这9个宏变量的取值即可。第5步、第6步、第8步和第9步中, 宏变量为字符型变量, 其取值需以英文单引号括起来, 尤其是第5步和第6步, 同时赋以因素的多个水平值, 此时每个水平都需以英文单引号括起来, 各水平之间需留有一个或多个空格。第7步中, 宏变量为数值型变量(由第11步中调用的 tabulate 过程决定, 必须为数值型变量), 水平值无须用英文单引号括起来, 但各水平之间同样需留有一个或多个空格。第10步是调用 plan 过程将数字进行随机化排列, 结果输出至数据集 sheji 中。第11步调用 tabulate 过程将设计方案以报表形式展现出来。

【主要输出结果及解释】

受试对象编号	剂量			
	阶段 1	阶段 2	阶段 3	阶段 4
受试对象 1	3	2	4	1
受试对象 2	4	1	2	3
受试对象 3	2	3	1	4
受试对象 4	1	4	3	2

第一行数据表示第1名受试对象在第1阶段至第4阶段依次接受剂量3、剂量2、剂量4和剂量1的某药物注射。其他可以此类推。

12.4.4 用 SAS 实现 2×2 交叉设计

【例 12-49】 某研究者为比较不同注射方式给药后胰岛素的吸收率, 选择 18 名健康志愿者进

行试验,计划采用交叉设计进行此试验。随机选取 9 名受试者在第一阶段采用喷射注射的方式接受 0.2IU/kg 体重的门冬胰岛素治疗,然后在第二阶段再采用传统胰岛素笔注射的方式接受 0.2IU/kg 体重的门冬胰岛素治疗,两个阶段之间设定一个充分的清洗期。另外 9 名受试者接受治疗的注射方式顺序与此相反,请给出具体的设计方案。

【分析】对 18 个健康志愿者进行编号,范围为 1~18。对注射方式进行编号,1 为喷射注射,2 为传统胰岛素笔注射。SAS 程序如下,程序名为 jcyx12_49.sas。

```
% let seed=10000;          /* 1 */
% let subject=18;          /* 2 */
% let time_values='第1阶段' '第2阶段'; /* 3 */
% let treat_values=1 2;    /* 4 */
% let subject_name='受试对象编号'; /* 5 */
% let treat_name='注射方式'; /* 6 */

proc plan seed=&seed;        /* 7 */
  factors subject=&subject time=
  2 ordered;
  treatments treat=2 cyclic;
  output out=a
    time cvals=(&time_values)
    treat nvals=(&treat_values);
run;

ods html style=analysis;   /* 8 */
proc tabulate style=[width=200]
  format=2.0;
  class subject time;
  var treat;
  table subject=' ', treat=&treat_name*
  time=' '*sum=' '/box=[label=&subject_name];
run;
ods html close;
```

【程序说明】第 1 步至第 6 步为设置宏变量,依次为随机种子数、受试对象个数、各阶段名称、各处理名称、受试对象名称(后加编号二字)、处理因素名称,使用时仅需根据实际情况修改这 6 个宏变量的取值即可。第 3 步、第 5 步和第 6 步中,宏变量为字符型变量,其取值需以英文单引号括起来,尤其是第 3 步,同时赋以因素的多个水平值,此时每个水平都需以英文单引号括起来,各水平之间需留有一个或多个空格。第 4 步中,宏变量为数值型变量(由第 8 步中调用的 tabulate 过程决定的,必须为数值型变量),水平值无须用英文单引号括起来,但各水平之间同样需留有一个或多个空格。第 7 步是调用 plan 过程将数字进行随机化排列,结果输出至数据集 sheji 中。第 8 步调用 tabulate 过程将设计方案以报表形式展现出来。

【主要输出结果及解释】

受试对象编号	注射方式	
	第 1 阶段	第 2 阶段
1	1	2
2	1	2
3	1	2
4	2	1
5	1	2
6	1	2
7	2	1
8	1	2
9	1	2
10	1	2
11	2	1
12	1	2

受试对象编号	注射方式	
	第 1 阶段	第 2 阶段
13	2	1
14	2	1
15	2	1
16	2	1
17	2	1
18	2	1

第一行数据表示 1 号志愿者在第一阶段接受 1 号注射方式(即喷射注射)处理,在第二阶段接受 2 号注射方式(即传统胰岛素笔注射)处理。其他可以此类推。

12.4.5 用 SAS 实现 3×3 交叉设计

【例 12-50】 某研究者欲研究从藤黄果中提取的羟基柠檬酸对摄食量的影响,选择 18 个年龄在 50~70 岁之间超重或肥胖的健康受试者,分别接受安慰剂,2800mg/d 羟基柠檬酸,5600mg/d 羟基柠檬酸的治疗。拟采用 3×3 交叉设计来安排此试验,采用随机化的方式确定每位受试对象接受三种处理的顺序,每种处理持续 6 周,结束后记录下一天的摄食量,然后经过 6 周的清洗期后,再接受下一种处理,请给出具体的设计方案。

【分析】 对 18 名健康受试者进行编号,范围为 1~18。对药物及剂量进行编号,1 为安慰剂(相当于 0mg/d 羟基柠檬酸),2 为 2800mg/d 羟基柠檬酸,3 为 2800mg/d 羟基柠檬酸。SAS 程序如下,程序名为 jcyx12_50.sas。

```
% let seed=10000;          /* 1 */
% let subject=18;           /* 2 */
% let time_values='第 1 阶段'
    '第 2 阶段' '第 3 阶段'; /* 3 */
% let treat_values=1 2 3; /* 4 */
% let subject_name='受试对象编号';
                                /* 5 */
% let treat_name='剂量';        /* 6 */
proc plan seed=&seed;           /* 7 */
    factors subject=&subject time=
3 ordered;
    treatments treat=3 cyclic;
    output out=a
    time cvals=(&time_values)
    treat nvals=(&treat_values);
run;

ods html style=analysis      /* 8 */
proc tabulate style=[width=200]
    format=2.0;
    class subject time;
    var treat;
    table subject=' ', treat=&treat
_name* time
    = ' ' * sum = ' '/box = [label =
&subject_name];
run;
ods html close;
```

【程序说明】 第 1 步至第 6 步为设置宏变量,依次为随机种子数、受试对象个数、各阶段名称、各处理名称、受试对象名称(后加编号二字)、处理因素名称,使用时仅需根据实际情况修改这 6 个宏变量的取值即可。第 3 步、第 5 步和第 6 步中,宏变量为字符型变量,其取值需以英文单引号括起来,尤其是第 3 步,同时赋以因素的多个水平值,此时每个水平都需以英文单引号括起来,各水平之间需留有一个或多个空格。第 4 步中,宏变量为数值型变量(由第 8 步中调用的 tabulate 过程决定的,必须为数值型变量),水平值无须用英文单引号括起来,但各水平之间同样需留有一个或多个空格。第 7 步是调用 plan 过程将数字进行随机化排列,结果输出至数据集 sheji 中。第 8 步调用 tabulate 过程将设计方案以报表形式展现出来。

【主要输出结果及解释】

受试对象编号	剂量		
	第 1 阶段	第 2 阶段	第 3 阶段
1	2	3	1
2	2	3	1
3	1	2	3
4	2	3	1
5	2	3	1
6	3	1	2
7	3	1	2
8	3	1	2
9	1	2	3
10	3	1	2
11	3	1	2
12	1	2	3
13	2	3	1
14	1	2	3
15	1	2	3
16	3	1	2
17	2	3	1

第一行数据表示 1 号受试者在第一阶段接受 2 号剂量药物(即 2800mg/d 羟基柠檬酸)处理,在第二阶段接受 3 号剂量药物(即 5600mg/d 羟基柠檬酸)处理,在第三阶段接受 1 号剂量药物(即安慰剂)处理。其他可以此类推。

12.4.6 用 SAS 实现析因设计

【例 12-51】 某研究者欲探讨 I 型多聚 ADP 核糖合成酶(PARP1)在心肌梗死后大鼠心肌组织中的表达及活性变化,选择雄性 Wistar 大鼠 36 只,210~260g,均结扎冠状动脉左前降支(LAD)近端建立急性心肌梗死(AMI)模型,并在术前 15~20min 及术后 2h 腹腔注射相应药物,以后每天注射同等剂量,直至处死当天为止。实验中涉及的药物有 30mg/kg 剂量的 3-氨基苯甲酰胺(3-AB)、100mg/kg 剂量的 3-氨基苯甲酰胺(3-AB)和 0.9% 氯化钠溶液(AMI 组),注射药物天数有 1 天、3 天和 7 天。拟采用析因设计安排此实验,请给出具体的试验设计方案。

【分析】 实验中涉及两个因素,分别是药物种类(30mg/kg 剂量的 3-AB、100mg/kg 剂量的 3-AB 和 0.9% 氯化钠溶液)和注射药物天数(1 天、3 天和 7 天),共可组合出 9 种实验条件,由于共有 36 只 Wistar 大鼠,所以每种实验条件下可进行 4 次独立重复实验(即 4 只大鼠)。对 36 只新西兰家兔进行编号,范围为 1~36。SAS 程序如下,程序名为 jcyx12_51.sas。

```

% let seed=10000;          /* 1 */      proc plan seed=&seed;          /* 12 */
% let subject=36;          /* 2 */      factors subject=&subject;
% let treat1=3;            /* 3 */      output out=b;
% let treat2=3;            /* 4 */      run;
% let repeat=4;            /* 5 */      data c;                      /* 13 */
% let treat1_values='30mg/kg 的 3-AB' merge a b;
    '100mg/kg 的 3-AB' '0.9% 氯化钠'; run;
                                /* 6 */      ods html style=analysis; /* 14 */

```

```
% let treat2_values = '1 天' '3 天'
'7 天'; /* 7 */
% let repeat_values = '重复试验 1'
'重复试验 2'
'重复试验 3' '重复试验 4'; /* 8 */
% let subject_name = 'Wistar 大鼠编号';
/* 9 */
% let treat_name = '药物种类 (左) 和
注射天数 (右)'; /* 10 */
proc plan seed = &seed; /* 11 */
factors treat1 = &treat1 treat2
= &treat1
repeat = &repeat;
output out = a
treat1 cvals = (&treat1_values)
treat2 cvals = (&treat2_values)
repeat cvals = (&repeat_values);
run;

proc tabulate data = c style =
[width=100]
format = 2.0;
class treat1 treat2 repeat;
var subject;
table treat1 = ' ' * treat2 = ' ',
subject =
&subject_name * repeat = ' ' * sum
= ' ' /
box = [label = &treat_name];
run;
ods html close;
```

【程序说明】 第1步至第10步为设置宏变量，依次为随机种子数、受试对象个数、试验因素1的水平数、试验因素2的水平数、独立重复试验的次数、试验因素1中各水平的名称、试验因素2中各水平的名称、各次独立重复试验的名称、受试对象名称(后加编号二字)、全部试验因素的总名称，使用时仅需根据实际情况修改这10个宏变量的取值即可。第6步至第10步中，宏变量为字符型变量，其取值需以英文单引号括起来，尤其是第6步至第8步中，同时赋以因素的多个水平值，此时每个水平都需以英文单引号括起来，各水平之间需留有一个或多个空格。第11步调用 plan 过程产生 treat1、treat2、repeat 三个变量各水平完全组合产生的36种条件，并对其进行随机化。第12步调用 plan 过程产生 subject 的编号，并对其进行随机化。第13步为数据步，实现对此前 plan 过程步产生的 a、b 两个数据集的合并。第14步是调用 tabulate 过程，以报表形式将设计方案展现出来。

【主要输出结果及解释】

药物种类 (左) 和注射天数 (右)		Wistar 大鼠编号			
		重复试验 1	重复试验 2	重复试验 3	重复试验 4
0.9% 氯化钠	1 天	12	25	16	14
	3 天	32	18	8	3
	7 天	4	24	28	22
100mg/kg 的 3- AB	1 天	29	30	34	5
	3 天	27	6	15	35
	7 天	23	36	11	33
30mg/kg 的 3- AB	1 天	20	2	9	13
	3 天	17	31	10	26
	7 天	7	19	21	1

第一行数据表示第6号、第31号、第5号及第23号新西兰家兔接受1nmol/L的胡椒碱处理。其他可以此类推。

12.4.7 用 SAS 实现含区组因素的析因设计

【例 12-52】 某研究者欲研究小鼠在以不同注射剂量(每天总注射量，因素 A)和不同注射频率(每天注射次数，因素 B)下药剂 ACTH 对尿总酸度的影响。其中，A 有3个水平：低剂量、中剂

量和高剂量; B 有 3 个水平: 1 次/天、2 次/天和 3 次/天。这两个因素共有 9 种水平组合, 称为 9 个处理组。研究者欲采用含区组因素的析因设计安排此试验, 先将 54 只小鼠按某些条件分为 6 个区组, 使每个区组中的 9 只小鼠在规定的方面条件最为接近, 然后, 将每个区组中的 9 只小鼠随机地分入 9 个处理组中。请给出具体的设计方案。

【分析】对 54 只小鼠进行编号, 区组 1 内的小鼠编号范围为 1~9, 区组 2 内的小鼠编号范围为 10~18, 区组 3 内的小鼠编号范围为 19~27, 区组 4 内的小鼠编号范围为 28~36, 区组 5 内的小鼠编号范围为 37~45, 区组 6 内的小鼠编号范围为 46~54。SAS 程序如下, 程序名为 jcyx12_52.sas。

```
% let seed=10000;          /* 1 */
% let block=6;              /* 2 */
% let treat1_level=3;        /* 3 */
% let treat2_level=3;        /* 4 */
% let treat1=低剂量@中剂量@高剂量; /* 5 */
% let treat2=1次/天@2次/天@3次/天; /* 6 */
% let subject_name='小鼠编号'; /* 7 */
% let block_name='区组';     /* 8 */
proc plan seed=&seed;         /* 9 */
  factors block=&block ordered
  condition=%eval(&treat1_level
* &treat1_level);
  output out=a;
run;
data b;                      /* 10 */
  set a;
  subject=_n_;
  do i=1 to &treat1_level;
    if condition<=i* &treat1_level
and condition>(i-1)* &treat1_level then do;
      a=i;
      b=condition-(i-1)* &treat1_level;
    end;
  end;
run;
% macro modify (treat1_level,
treat2_level);               /* 11 */
% do i=1 %to &treat1_level;
data b;
  set b;
  if a=&i then a1="%sysfunc(scan
(&treat1, &i, @), 20)";
run;
% end;
% do i=1 %to &treat2_level;
data b;
  set b;
  if b=&i then
  b1="%sysfunc(scan(&treat2, &i,
@), 20)";
run;
% end;
% mend modify;
% modify (&treat1_level, &treat2_level); /* 12 */
ods html style=analysis;
/* 13 */
proc tabulate style=[width=60]
format=2.0;
  class block a1 b1;
  var subject;
  table block=' ', subject=
&subject_name*a1
=' '*b1=' '*sum=' '/box=[label=
&block_name];
run;
ods html close;
```

【程序说明】第 1 步至第 8 步为设置宏变量, 依次为随机种子数、区组因素水平数、试验因素 1 的水平数、试验因素 2 的水平数、试验因素 1 中各水平的名称(各水平间以 @ 符号间隔, 方便后续调用)、试验因素 2 中各水平的名称(各水平间以 @ 符号间隔, 方便后续调用)、受试对象名称(后加编号二字)、区组因素名称, 使用时仅需根据实际情况修改这 8 个宏变量的取值即可。第 9 步调用 plan 过程产生初步的设计方案, 并对其进行随机化。第 10 步为数据步, 对此前 plan 过程步产生的数据集 a 进行操作, 目的是产生小鼠编号, 并将实验条件还原为各因素的水平组合。第 11 步为创建宏程序, 从而使以数字形式给出的实验因素的水平转换为第 5 步、第 6 步中的字符, 使最终的试验方案更容易理解。第 12 步为调用之前创建的宏程序 modify。第 13 步是调用 tabulate 过程, 以报表形式将设计方案展现出来。

【主要输出结果及解释】

区组	小鼠编号								
	低剂量			高剂量			中剂量		
	1 次/天	2 次/天	3 次/天	1 次/天	2 次/天	3 次/天	1 次/天	2 次/天	3 次/天
1	7	4	6	8	9	2	3	5	1
2	15	18	11	12	17	13	14	16	10
3	26	20	24	22	27	25	23	21	19
4	33	36	30	31	29	35	34	32	28
5	38	39	43	45	44	40	37	42	41
6	52	49	54	53	46	48	50	47	51

第一行数据表示 7、4、6 号小鼠分别接受 1 次/天的低剂量、2 次/天的低剂量、3 次/天的低剂量药剂 ACTH 注射，8、9、2 号小鼠分别接受 1 次/天的高剂量、2 次/天的高剂量、3 次/天的高剂量药剂 ACTH 注射，3、5、1 号小鼠分别接受 1 次/天的中剂量、2 次/天的中剂量、3 次/天的中剂量药剂 ACTH 注射。其他可以此类推。

12.4.8 用 SAS 实现平衡不完全区组设计

【例 12-53】 某研究者欲比较生陈皮及采用 8 种不同炮制方法(去白、土炒、清炒、糠炒、麸炒、甘草炙、盐炙、福州制)得到的炮制陈皮对气管和回肠平滑肌的抑制作用，选择 18 窝豚鼠为受试对象，由于每窝仅能保证 5 只，拟采用平衡不完全区组设计安排此试验，请给出具体的设计方案。

【分析】 对受试对象进行编号，第 1 窝内的受试对象编号范围为 1~5，第 2 窝内的受试对象编号范围为 6~10，以此类推，直至第 18 窝，其受试对象编号范围为 86~90。生陈皮及去白、土炒、清炒、糠炒、麸炒、甘草炙、盐炙、福州制等 9 种炮制陈皮依次编为陈皮 1~陈皮 9。SAS 程序如下，程序名为 jcyx12_53.sas。

```
% let seed=10000; /* 1 */
% let block=18; /* 2 */
% let treat_level=9; /* 3 */
% let subject_per_block=5; /* 4 */
% let subject_name='豚鼠编号'; /* 5 */
% let block_name='区组'; /* 6 */
% let treat_name=生陈皮@ 去白@ 土炒
@ 清炒@ 糠炒@ 麸炒@ 甘草炙@ 盐炙@ 福
州制; /* 7 */
% let length=8; /* 8 */
data a; /* 9 */
  length treatment $ &length;
  input treatment;
  cards;
run;
% macro modify(treat_level); /* 10 */
% do i=1 % to &treat_level;
data b;
  treatment="% sysfunc(scan(&treat_
name, &i, @ ), &length)";
run;
data a;
  set a b;
run;
% end;
% mend modify;

% modify(&treat_level); /* 11 */
proc optex data=a seed=&seed cod-
ing=orth; /* 12 */
  class treatment;
  model treatment;
  blocks structure=(&block)
  &subject_per_block;
  output out=c;
run;
data d; /* 13 */
  set c;
  subject=_n_;
run;
ods html style=analysis; /* 14 */
proc tabulate style=[width=70]
format=2.0;
  class block treatment;
  var subject;
  table
  block=' ', subject=&subject_name
  * treatment=' '* sum=' '/box=[la-
bel=&block_name];
run;
ods html close;
```


【程序说明】 第 1 步至第 8 步为设置宏变量,依次为随机种子数、区组因素水平数、处理因素的水平数、每个区组中的受试对象数、受试对象名称(后加编号二字)、区组因素名称、处理因素各水平的名称(各水平间以@ 符号间隔,方便后续调用)和处理因素各水平中取值的最大字节数(每个英文字母或数字为 1 个字节,每个汉字为 2 个字节),使用时仅需根据实际情况修改这 8 个宏变量的取值即可。第 9 步为建立一个包含 treatment 变量的空数据集 a。第 10 步为创建宏程序,从而将处理因素的各水平以列的形式存入数据集 a 中。第 11 步为调用之前创建的宏程序 modify。第 12 步调用 optex 过程产生试验设计方案,并对其进行随机化。第 13 步对数据集 c 进行操作,产生受试对象编号,相当于为每个受试对象分组。第 14 步是调用 tabulate 过程,以报表形式将设计方案展现出来。

【主要输出结果及解释】

区组	豚鼠编号								
	福州制	甘草炙	糠炒	清炒	去白	生陈皮	土炒	盐炙	麸炒
1	4	—	—	3	5	—	1	—	2
2	—	7	—	—	—	10	8	6	9
3	15	12	11	—	—	13	14	—	—
4	20	—	17	—	16	—	—	18	19
5	—	22	25	21	—	—	—	23	24
6	28	—	—	—	30	27	26	29	—
7	32	35	34	—	—	33	—	—	31
8	—	38	39	37	40	—	36	—	—
9	45	—	41	44	—	—	42	43	—
10	—	—	50	46	47	49	—	—	48
11	—	51	55	—	52	—	54	53	—
12	60	58	—	57	—	—	56	—	59
13	—	65	—	64	62	63	—	61	—
14	70	—	67	66	—	69	—	68	—
15	—	—	73	—	72	75	74	—	71
16	78	79	—	—	80	—	—	77	76
17	85	83	—	81	84	82	—	—	—
18	—	—	—	88	—	90	89	87	86

第一行数据表示第 1 号~第 5 号受试对象分别接受土炒陈皮、麸炒陈皮、清炒陈皮、福州制陈皮和去白陈皮的处理。其他可以此类推。

参 考 文 献

- [1] 胡良平. 统计学三型理论在实验设计中的应用. 北京: 人民军医出版社, 2006, 20-214.
- [2] 刘玉秀, 姚晨, 杨友春等. 随机化临床试验及随机化的 SAS 实现. 中国临床药理学与治疗学, 2001; 6(3): 193-195.
- [3] 薛富波, 张文彤, 田晓燕. SAS 8.2 统计应用教程. 北京: 兵器工业出版社, 北京希望电子出版社, 2004: 20-216.
- [4] SAS Institute Inc. SAS/STAT 9.2 User's Guide. North Carolina: SAS Institute Inc. 2008: 4725-4757.
- [5] SAS Institute Inc. SAS/QC 9.2 User's Guide. North Carolina: SAS Institute Inc. 2008: 977-1054.
- [6] SAS Institute Inc. Base 9.2 Procedures Guide. North Carolina: SAS Institute Inc. 2008: 1297-1054.
- [7] 胡良平. 科研设计与统计分析. 北京: 军事医学科学出版社, 2012: 129-278.
- [8] 胡良平. SAS 统计分析教程. 北京: 电子工业出版社, 2010: 674-706.

第 13 章 统计表达与描述的关键技术

统计表与统计图都是表达统计资料的重要工具。一般来说，统计表表达统计资料具体、明确；而统计图表达统计资料形象、生动。统计表可用来表达定性资料、定量资料和混合型资料；而统计图一般只适合表达定性资料与定量资料，但不便于同时表达定量与定性混合型资料。本章主要介绍如何用统计表与统计图表达各种统计资料。

13.1 以简化形式表达多组定量资料的关键技术

所谓“以简化形式表达多组定量资料的统计表”，就是对表达某种试验设计定量资料原始数据的表中，用“ $\bar{x} \pm s$ ”的形式表达各不同条件下服从对称分布的重复试验数据的平均值与标准差；用“ $M(Q_1 \sim Q_3)$ ”或“ $M(Q_R = Q_1 \sim Q_3)$ ”的形式表达各不同条件下呈非对称分布的重复试验数据的中位数与下四分位数和上四分位数（或中位数与四分位数间距）。也就是说，在一切试验设计下的定量资料，既可以用原始数据形式表达，也可用简化形式表达，为节约篇幅，本节仅举两个例子。

13.1.1 用“ $\bar{x} \pm s$ ”的形式表达析因设计定量资料

表 13-1 各组大鼠术后第 1、3、7 天左室收缩末期内径大小 (n=4)

3-AB 注射量 (mg/kg)	左室收缩末期内径 ($\bar{x} \pm s$, cm)		
	* : 1	3	7
0	4.10 \pm 0.50	3.90 \pm 0.39	4.10 \pm 0.22
30	2.40 \pm 0.20	2.60 \pm 0.21	3.70 \pm 0.29
100	2.50 \pm 0.29	3.10 \pm 0.61	3.30 \pm 0.20

注：* 代表“处死时间(d)”

表 13-2 不同人群血清胱抑素 C 和 Cr 结果比较 ($\bar{x} \pm s$)

肾功能状况	血 压 高 否	性 别	例数 (人)	血清胱抑素 C (mg/L)	Cr (mg/L)
正常	否	男	111	0.98 \pm 0.20	74.7 \pm 11.8
		女	97	0.97 \pm 0.18	61.3 \pm 8.4
	是	男	141	1.00 \pm 0.21	75.4 \pm 11.4
		女	63	1.00 \pm 0.27	61.5 \pm 8.8
早期损伤	否	男	70	1.02 \pm 0.22	76.6 \pm 12.0
		女	39	1.01 \pm 0.18	60.5 \pm 10.8
	是	男	94	1.06 \pm 0.23	79.6 \pm 10.8
		女	30	1.07 \pm 0.37	64.1 \pm 10.3

13.1.2 用“ $M(Q_1 \sim Q_3)$ ”的形式表达裂区设计定量资料

表 13-3 不同瘤株与不同浓度的蛇毒共同作用后对小鼠抑瘤效果的影响 ($n=3$)

瘤株 A	瘤重 [$M(Q_1 \sim Q_3)$, g]			
	* : B_1	B_2	B_3	B_4
A_1 (S180)	0.76 (0.36 ~ 0.80)	0.31 (0.26 ~ 0.36)	0.17 (0.16 ~ 0.28)	0.13 (0.11 ~ 0.28)
A_2 (HS)	0.57 (0.43 ~ 0.74)	0.46 (0.32 ~ 0.50)	0.30 (0.20 ~ 0.42)	0.32 (0.26 ~ 0.36)
A_3 (EC)	0.32 (0.31 ~ 0.55)	0.20 (0.15 ~ 0.24)	0.26 (0.18 ~ 0.38)	0.21 (0.14 ~ 0.25)
A_4 (ARS)	0.48 (0.33 ~ 0.57)	0.29 (0.18 ~ 0.30)	0.28 (0.27 ~ 0.44)	0.30 (0.22 ~ 0.37)

注: * 代表“蛇毒浓度(B)”

13.2 编制常用统计表的关键技术

13.2.1 表达定性资料的统计表

1. 横断面研究设计的 2×2 表资料

对一组受试者进行某项问题的调查或进行某项试验研究,在短期内可按两个水平分别就原因和结果变量对受试对象进行分组,从而获得的 2×2 表资料,称为来自横断面研究设计的 2×2 表资料,见表 13-4。

2. 队列研究设计的 2×2 表资料

依据专业知识,提出可能的危险因素,按此危险因素“接触、不接触”将全部受试者分成两组,相当于将他们排成两个队伍,研究者对此两个队列人群进行随访追踪观察若干年,最后记录各组中的受试者患某病与不患该病的人数,这样收集到的 2×2 表资料称为来自队列研究设计的 2×2 表资料,见表 13-5。

表 13-4 两组工人高血压患病情况

工作环境	例数(人)		
	* : 是	否	合计
高温作业	171	788	959
非高温作业	27	230	257
合计	198	1018	1216

注: * 代表“是否患高血压”

表 13-5 两种治疗方法随访 3 年
大肠癌的复发转移情况

治疗方法	例数(人)		
	* : 是	否	合计
放化疗 + 中药	17	67	84
放化疗	38	40	78
合计	55	107	162

注: * 代表“是否复发转移”

3. 病例-对照研究设计的 2×2 表资料

为什么有些人会患某种病,而另一些人却没患该病?研究者依据专业知识,提出导致该病的可疑危险因素,寻找与患者条件相近的另一组人作为正常对照者,对他们进行回顾性调查,并按此可疑危险因素的两个水平“接触、不接触”将两组被调查者分别再分成两组,并记录各组中的接触与不接触者的人数,这样收集到的 2×2 表资料称为来自病例-对照研究设计的 2×2 表资料,见表 13-6。

4. 配对研究设计的 2×2 表资料

对性质相同的一组样品同时用两种检测方法进行检测,每种检测方法检测的结果都分为“阳性(+)”与“阴性(-)”,数出两种检测方法同时判定为阳性、阴性的样品数以及检测结果不一致的样品数,将结果列成配对研究设计的 2×2 表资料格式,见表 13-7。

表 13-6 是否患乳腺癌与有无人工流产史情况

有无人工 流产史	例数(人)		
	* :是	否	合计
有	1003	1075	2078
无	492	498	990
合计	1495	1573	3068

注：* 代表“是否患乳腺癌”

表 13-7 脓毒症 BNP 与血瘀证金标准诊断结果比较

BNP 法	例数(人)		
	* :血瘀证	非血瘀证	合计
血瘀证	86	35	121
非血瘀证	9	44	53
合计	95	79	174

注：* 代表“金标准”

5. 原因变量为二值变量而结果变量为多值名义变量的 $2 \times C$ 表资料

这种列联表可视为双向无序的 $R \times C$ 表资料的特例，见表 13-8。

表 13-8 维吾尔族与回族居民血型的频数分布

民 族	例数(人)				
	血型：A	B	O	AB	合 计
维吾尔族	442	483	416	172	1513
回族	369	384	487	115	1355
合计	811	867	903	287	2868

6. 原因变量为二值变量而结果变量为多值有序变量的 $2 \times C$ 表资料

两种药物治疗心绞痛 4 周后的疗效观察，见表 13-9。

表 13-9 两种药物治疗心绞痛的疗效

治 疗 药 物	例数(人)			
	疗效：优	显效	无效	合 计
益气通脉口服液	58	43	9	110
舒心口服液	32	61	16	109
合计	90	104	25	219

7. 原因变量为多值名义变量而结果变量为二值变量的 $R \times 2$ 表资料

这种列联表可视为双向无序的 $R \times C$ 表的特例，见表 13-10。

表 13-10 脂肪肝中医证型与 CT 分型的关系

中 医 证 型	例数(人)		
	CT 分型：弥漫性脂肪肝	局灶型脂肪浸润	合 计
肝郁脾虚	34	8	42
痰瘀互结	5	1	6
痰湿内阻	13	4	17
肝肾不足	7	5	12
湿热内蕴	28	8	36
合计	87	26	113

8. 原因变量为多值有序变量而结果变量为二值变量的 $R \times 2$ 表资料

让雌性小鼠暴露于三种不同剂量的二氯乙烷之下, 经过一段时间后, 观察各剂量组小鼠得肺癌的情况, 试验结果见表 13-11。

表 13-11 138 只雌性小鼠暴露在三种不同剂量的二氯乙烷之下肺癌发生情况

二氯乙烷剂量 (mg/kg)	例数(只)		
	患肺癌	未患肺癌	合 计
0	2	38	40
1	7	43	50
2	15	33	48
合计	24	114	138

9. 双向无序的 $R \times C$ 表资料

双向无序的 $R \times C$ 表资料, 顾名思义, 就是表中两个定性变量(在数量关系上来说)都是名义变量, 见表 13-12。

表 13-12 某医院三年间四种甲状腺疾病在四季中发病人数的分布情况

甲状腺 病分类	例数(人)				
	季节: 春	夏	秋	冬	合 计
甲亢	411	451	294	284	1440
亚甲炎	249	329	331	204	1113
甲低	60	61	59	52	232
甲状腺肿瘤	45	50	46	40	181
合计	765	891	730	580	2966

10. 结果变量为有序变量的单向有序的 $R \times C$ 表资料

单向有序的 $R \times C$ 表资料, 要特别强调结果变量是有序的, 见表 13-13; 若仅原因变量是有序的单向有序的 $R \times C$ 表资料, 仍应将其视为“双向无序的 $R \times C$ 表资料”, 见表 13-14(认为“HBV 感染结局”是有序变量)。

表 13-13 术后 15d 前胃癌小鼠肺转移情况

组 别	例数(只)				
	肺转移程度:0	I	II	III	合 计
模型组	1	3	5	9	18
5-FU 化疗组	5	5	5	4	19
健脾益肾方治疗组	6	5	6	2	19
合计	12	13	16	15	56

表 13-14 三组患者 VDR-Fok I 基因型分布比较

HBV 感染结局	例数(人)			
	基因型: TT	CT	CC	合 计
慢性乙肝患者	120	184	87	391
HBV 携带者	77	122	45	244
HBV 自限性感染者	87	88	37	212
合计	284	394	169	847

11. 双向有序且属性不同的 $R \times C$ 表资料

当 $R \times C$ 列联表中的两个定性变量都是有序变量，且它们的属性不同，此时，称这样的列联表资料为双向有序且属性不同的 $R \times C$ 列联表资料，见表 13-15。

表 13-15 三组患者术后疗效比较

手术时间	例数(人)				
	疗效: 优	良	可	差	合 计
伤后 8h 内	17	12	2	0	31
伤后 8h ~ 7d	5	6	5	1	17
伤后 7 ~ 14d	12	14	2	0	28
合计	34	32	9	1	76

12. 双向有序且属性相同的 $R \times C$ 表资料

当 $R \times C$ 列联表中的两个定性变量都是有序变量，且它们的属性相同，此时，称这样的列联表资料为双向有序且属性相同的 $R \times C$ 列联表资料，见表 13-16。

表 13-16 两种诊断方法对 117 例直肠癌患者浸润深度的诊断结果

腔内超声 诊断分级	例数(人)				
	* :T ₁	T ₂	T ₃	T ₄	合 计
T ₁	7	1	0	0	8
T ₂	1	15	10	0	26
T ₃	0	13	60	2	75
T ₄	0	0	0	8	8
合计	8	29	70	10	117

注: * 代表“手术病理诊断分级”

13. 结果变量为二值变量的三维列联表资料

表 13-17 中包含两个原因变量(不同医院、不同程度静脉炎)与一个结果变量(疗效)。

表 13-17 如意金黄散外涂治疗不同程度的化疗药物性静脉炎及其疗效

不 同 医 院	不同程度静脉炎	例数(人)	
		疗效: 有效	无效
甲医院	I 级	17	0
	II 级	40	0
	III 级	46	2
	IV 级	12	3
乙医院	I 级	36	10
	II 级	18	10
	III 级	24	12
	IV 级	8	12

14. 结果变量为多值有序变量的高维列联表资料

表 13-18 是结果变量为多值有序变量的三维列联表资料。

表 13-18 中药贴剂是否加磁治疗乳腺增生症的治疗效果

增生症类型	磁粉用否	例数(人)			
		疗效:痊愈	显效	有效	无效
乳腺组织增生	否	0	2	4	12
	用	9	30	7	0
乳腺腺病	否	0	1	3	24
	用	4	48	15	5
乳腺囊肿病	否	0	0	4	10
	用	2	13	8	9

15. 结果变量为多值名义变量的高维列联表资料

表 13-19 是结果变量为多值名义变量的三维列联表资料。

表 13-19 不同性别不同中医证型患者间 Megsin 基因 E1-5' UTR 区 3 个 SNP 的分布差异

性 别	中 医 证 型	例数(人)		
		SNP: GG	GA	AA
男	气阴两虚	28	4	0
	肝肾阴虚	10	11	1
女	气阴两虚	35	5	0
	肝肾阴虚	10	14	2

16. 具有重复测量设计的定性资料

选择符合纳入标准的急性上呼吸道感染患者 339 例,随机分为三组,分别接受一清双解胶囊(治疗组 A)、一清双解片剂(治疗组 B)和柴黄片(对照组 C)治疗。观察三组患者治疗前后的发热恶寒程度,试验结果见表 13-20。

表 13-20 三种急性上呼吸道感染患者治疗前后发热恶寒程度比较

药 物 种 类	发热恶寒程度		例数(人)
	时间: 治疗前	治疗后	
一清双解胶囊	1	1	n_1
	1	2	n_2
	\vdots	\vdots	\vdots
	4	4	n_{16}
一清双解片剂	1	1	n_{17}
	1	2	n_{18}
	\vdots	\vdots	\vdots
	4	4	n_{32}
柴黄片	1	1	n_{33}
	1	2	n_{34}
	\vdots	\vdots	\vdots
	4	4	n_{48}

表体内“发热恶寒程度”之下 1、2、3、4 分别代表 -、+、++、+++，例数列给出各行的例数。

13.2.2 表达原始定量资料的统计表

1. 随机区组设计一元定量资料原始数据(表 13-21)

表 13-21 不同药物作用后小白鼠肉瘤重量

区 组	肉瘤重量(g)		
	药物种类: A	B	C
1	0.82	0.65	0.51
2	0.73	0.54	0.23
3	0.43	0.34	0.28
4	0.41	0.21	0.31
5	0.68	0.43	0.24

2. 拉丁方设计一元定量资料原始数据(表 13-22)

表 13-22 伴有三柱损伤椎体压缩骨折在腰椎旋转 2° 负荷下的角度变化

受 试 者 号	固定器械种类及角度变化(度)		
	测定顺序: 1	2	3
1	C(0.50)	B(0.48)	A(0.11)
2	A(0.10)	C(0.39)	B(0.33)
3	B(0.55)	A(0.48)	C(0.56)

注: A、B、C 分别代表 AF 钉固定、Dick 钉固定及 ARRIF 钉固定。

3. 3 × 3 交叉设计一元定量资料原始数据(表 13-23)

表 13-23 18 名受试者使用不同制剂的 C_{\max} 结果

受 试 者 号	组 别	制剂种类与 C_{\max} (mg/L)		
		时期: 1	2	3
7	1	A(99.75)	B(71.73)	R(91.89)
12	1	A(68.48)	B(86.04)	R(81.39)
18	1	A(54.63)	B(36.86)	R(54.84)
5	2	A(38.47)	R(55.18)	B(26.30)
6	2	A(54.70)	R(55.58)	B(35.11)
13	2	A(31.13)	R(26.04)	B(57.05)
1	3	B(47.73)	R(37.43)	A(42.85)
4	3	B(49.42)	R(28.56)	A(61.49)
9	3	B(45.56)	R(46.13)	A(53.41)
2	4	B(59.75)	A(28.90)	R(54.04)
11	4	B(24.61)	A(32.11)	R(22.69)
15	4	B(16.43)	A(18.08)	R(21.42)
8	5	R(49.30)	A(45.17)	B(44.60)

续表

受试者号	组别	制剂种类与 C_{\max} (mg/L)		
		时期: 1	2	3
10	5	R(73.75)	A(66.78)	B(85.71)
17	5	R(46.86)	A(27.96)	B(20.26)
3	6	R(82.15)	B(75.52)	A(81.54)
14	6	R(52.58)	B(50.05)	A(54.46)
16	6	R(38.64)	B(53.35)	A(55.68)

注: A、B、R 分别代表氯雷他定片剂、氯雷他定胶囊剂和氯雷他定参比制剂。

13.3 绘制用于直观判断定量变量分布情况的统计图的关键技术

13.3.1 直方图 (Histogram) 应用场合及实例

适于表达一组定量资料的频数分布情况。其作用就是把用频数分布表表达的资料用图形方式显示出来,使各组段上的频数分布情况一目了然。它是用各组段上小长方形的面积来表示频数。因此,绘图时,纵轴上数据的单位应理解为(1/组距)。这样一来,各组段上以“1/组距”为单位的频数乘以组距,在数值上正好等于该组段上的频数。既保证了绘图上的方便,又保证了直方图用“面积”表示频数的要求。

【例 13-1】 某储蓄所抽查了 50 名顾客办理业务的等待时间(分),数据如下: 22, 14, 33, 17, 24, 17, 12, 14, 24, 11, 15, 20, 18, 3, 25, 24, 34, 37, 42, 35, 30, 23, 34, 22, 13, 34, 8, 22, 31, 24, 25, 0, 1, 21, 28, 33, 32, 23, 14, 4, 31, 42, 34, 26, 14, 25, 40, 14, 42, 6。试选用合适的统计图类型将其表达出来。

【分析与解答】 这个问题实际上是要求使用统计图表示一组定量资料的频数分布情况,显然,选择直方图比较合适,见图 13-1。

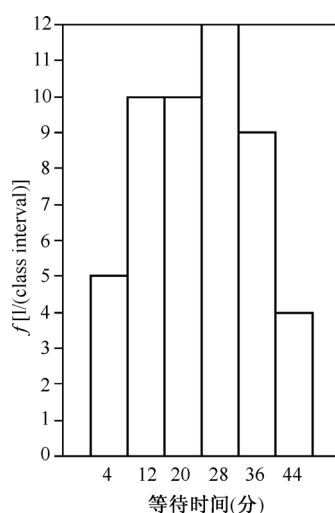


图 13-1 直方图的式样

绘制图 13-1 的 SAS 程序如下,程序名为 jcyx13_1.sas。

```
datajcyx13_1;
  input time@@ ;
  cards;
22 14 33 17 24 17 12 14
24 11 15 20 18 3 25 24
34 37 42 35 30 23 34 22
13 34 8 22 31 24 25 0
1 21 28 33 32 23 14 4
31 42 34 26 14 25 40 14
42 6
;
run;
goption hsize=4 vsize=4;
pattern1 value=empty color=blue;
axis1 label = (justify = center '等待时间(分)')
  minor=none major=none
  offset = (5, 5);
axis2 label = ( angle = 90 font = italic
  'f (1/(class interval))')
  minor=none major=(height=0.8)
  offset = (0, 0);

ods html;
proc gchart data=jcyx13_1;
  vbar time/type=freq maxis=axis1
  midpoints=4 to 44 by 8
  raxis=axis2 space=0 width=4;
run;
ods html close;
```

13.3.2 P-P 图、Q-Q 图及概率图应用场合及实例

用于反映定量观测指标是否服从某特定概率分布，如正态分布、对数正态分布、威布尔分布等。

P-P 图 (Probability-Probability Plot)：将定量指标的观测值按由小到大的顺序排列，然后计算其实际累积频率，以实际累积频率对被检验分布（如正态分布）的理论累积频率作图，称为 P-P 图。若原资料确实服从假定的概率分布，那么，图形大体上将呈直线。

Q-Q 图 (Quantile-Quantile Plot)：将定量指标的观测值按由小到大的顺序排列，以实际观测值对被检验分布（例如正态分布）的理论分位数作图，称为 Q-Q 图。若原资料确实服从假定的概率分布，那么，图形大体上将呈直线。

概率图 (Probability Plot)：将定量指标的观测值按由小到大的顺序排列，以实际观测值对被检验分布（如正态分布）的理论百分位数作图，称为概率图。若原资料确实服从假定的概率分布，那么，图形大体上将呈直线。

在实际应用中，切记不能将三者混淆。P-P 图是将实际累积频率与理论累积频率作图，而概率图是将一组排序后的观测数据与某一特定分布的理论百分位数比较，Q-Q 图则涉及观测数据与理论分位数。Q-Q 图与 P-P 图相比较，Q-Q 图效率较高。概率图与 Q-Q 图较为相似，需要注意的是，一个是百分位数，另一个则是分位数，Q-Q 图在分布参数的图形估计方面要优于概率图，而概率图可以更好地表示出百分位数。

【例 13-2】某工厂欲研究钢板切割中两洞的长度是否服从正态分布，选取 50 个切割好的钢板进行试验，测量每块钢板上两洞之间的距离 (cm)，数据如下：9.80, 10.20, 10.27, 9.70, 9.76, 10.11, 10.24, 10.20, 10.24, 9.63, 9.99, 9.78, 10.10, 10.21, 10.00, 9.96, 9.79, 10.08, 9.79, 10.06, 10.10, 9.95, 9.84, 10.11, 9.93, 10.56, 10.47, 9.42, 10.44, 10.16, 10.11, 10.36, 9.94, 9.77, 9.36, 9.89, 9.62, 10.05, 9.72, 9.82, 9.99, 10.16, 10.58, 10.70, 9.54, 10.31, 10.07, 10.33, 9.98, 10.15。试选用合适的统计图类型将下面的资料表示出来。

【分析与解答】 这是考察一组定量观测数据是否服从某种特定分布(此处为考察是否服从正态分布),可以选用 P-P 图、Q-Q 图或概率图,本数据为钢板上两洞之间的距离,选择正态分布进行作图,具体见图 13-2、图 13-3 和图 13-4。

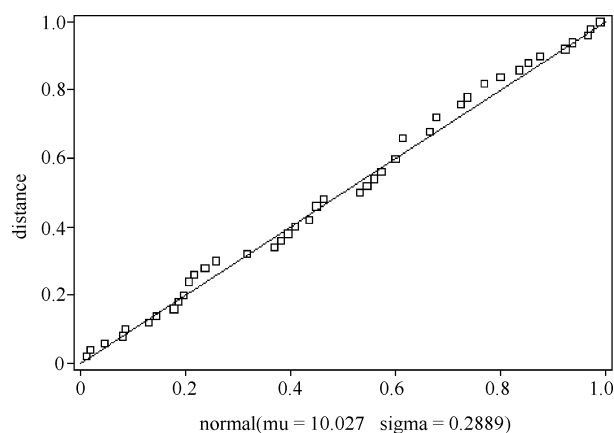


图 13-2 P-P 图的式样

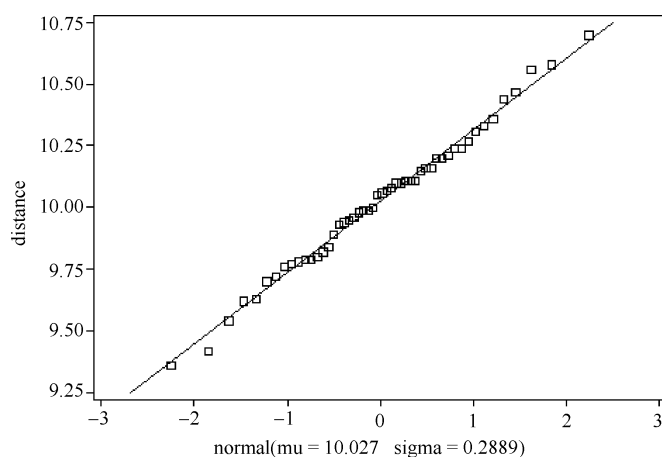


图 13-3 Q-Q 图的式样

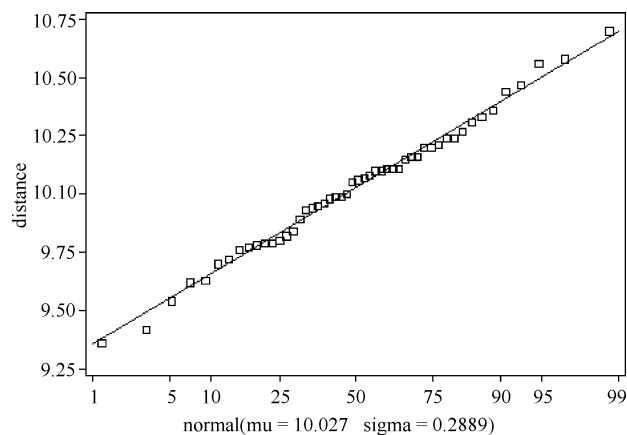


图 13-4 概率图的式样

绘制图 13-2、图 13-3 和图 13-4 的 SAS 程序如下，程序名为 jcyx13_2. sas。

```
datajcyx13_2;
  input distance@@ ;
  cards;

  9.80 10.20 10.27 9.70 9.76 10.11
10.24 10.20 10.24 9.63 9.99 9.78
10.10 10.21 10.00 9.96 9.79 10.08
9.79 10.06 10.10 9.95 9.84 10.11
9.93 10.56 10.47 9.42 10.44 10.16
10.11 10.36 9.94 9.77 9.36 9.89
9.62 10.05 9.72 9.82 9.99 10.16
10.58 10.70 9.54 10.31 10.07 10.33
9.98 10.15
;
run;
symbol v=square height=2.5pct;
goption hsize=5 vsize=4;
axis1 label = (angle = 90 font = i-
talic);
axis2 label = (font = italic);

ods html;
proc capability data = jcyx13_2 no-
print;
  ppplot distance/normal (mu = est
sigma = est)
    vaxis = axis1 haxis = axis2;
  qqplot distance/ normal (mu = est
sigma = est)
    vaxis = axis1 haxis = axis2 no-
legend;
  probplot distance/ normal (mu =
est sigma = est)
    vaxis = axis1 haxis = axis2 no-
legend;
run;
ods html close;
```

参 考 文 献

- [1] 胡良平. 科研设计与统计分析. 北京: 军事医学科学出版社, 2012: 27-70.
- [2] 胡良平. SAS 统计分析教程. 北京: 电子工业出版社, 2010: 656-673.

第 14 章 统计分析方法合理选择的关键技术

14.1 合理选择统计分析方法的一般步骤

在进行统计分析方法的选择时,需要考虑以下几个关键点:分析目的、设计类型、比较类型、结果变量性质及同时考察的个数、资料所具备的前提条件。具体来说,选择方法时,一般步骤如下。

(1)紧扣分析目的。统计分析方法一般可以分为定量或定性资料的统计分析方法、相关和回归分析方法、多因素分析方法、多元统计分析方法等几大类。在选择分析方法之前,需要先对分析目的有一个整体的认识:若是研究不同组间定量或定性指标有无统计学差异,可采用相应设计类型定量或定性资料的差异性分析方法;若是研究变量之间的相互和依赖关系,可采用相关与回归分析;若是研究多个原因对一个结果变量的影响,一般采用多因素分析;若研究一个或多个原因同时对多个定量结果变量的影响,则一般选择多元统计分析。当然,同一大类统计分析方法,也能细化成多种,实际选择分析方法时,还要根据具体的分析目的来确定。例如,对双向有序且属性不同的 $R \times C$ 列联表资料而言,比较 R 个水平组之间在结果变量各水平上的频数分布有无统计学差异可选择 χ^2 检验或 Fisher 精确检验;比较 R 个水平组之间在结果变量上的整体情况有无统计学差异,可选择秩和检验或 Ridit 分析;要研究原因变量与结果变量之间有无相关性,可选择 Spearman 秩相关分析或典型相关分析;要研究结果变量是否随原因变量的变化呈现出某种线性变化趋势,可选择线性趋势检验。再如,对单组设计多元定量资料而言,单纯研究其中两个连续型随机变量之间的线性相关强度,可选择简单相关分析或 Spearman 秩相关分析;研究当其他随机变量固定不变时或者消除了其他随机变量的影响后,两个随机变量之间的线性相关强度,可选择偏相关分析;研究某个随机变量与其他多个随机变量之间的线性相关强度,可选择复相关分析;研究某两组随机变量之间的线性相关强度,可选择典型相关分析。

(2)认清设计类型。不同设计类型的资料,都有其对应的统计分析方法。实际处理数据时,切记不能无视设计类型,用 t 检验或 χ^2 检验等方法作为万能钥匙来处理一切资料。对于定量资料而言,需要辨清资料的具体设计类型,看它到底是单组设计、成组设计、配对设计、单因素多水平设计、随机区组设计、析因设计等中的哪一种。对于定性资料,虽常以列联表形式给出,但其中也隐藏着“设计类型”,如通常的二维列联表资料可能是单因素设计定性资料(可分为成组设计与单因素多水平设计)或配对设计(含配对设计扩大形式)定性资料,而三维列联表资料通常为两因素析因设计定性资料。当然,分析定性资料时,在识别出设计类型的基础上,还要知道具体的列联表类型,这样更有利于正确选择统计分析方法。

(3)认准列联表类型。如何才能正确分析定性资料呢?认准列联表的类型至关重要。通常,定性资料可用二维列联表和高维列联表呈现出来。二维列联表可划分为 2×2 列联表、 $2 \times C$ 列联表、 $R \times 2$ 列联表、 $R \times C$ 列联表四种类型;高维列联表可划分为结果变量为二值、多值名义和多值有序三种类型。 2×2 列联表还可依据研究类型被进一步划分为横断面研究设计、队列研究设计、病例-对照研究设计、配对研究设计的 2×2 列联表四个亚型; $R \times C$ 列联表还可依据表中两个定性变量的

特性被进一步划分为双向无序、仅结果变量有序的单向有序、双向有序且属性不同和双向有序且属性相同的 $R \times C$ 列联表四个亚型。面对每一个列联表资料，当认准后，还要结合分析目的、自变量数目的多少，才能选择合适的统计分析方法。

(4) 选好比较类型。临床试验中的比较类型，按照研究目的可分为差异性试验、优效性试验、等效性试验和非劣效性试验，所以在处理相应资料时，一定要结合研究目的确定好比较的类型。差异性试验主要用来显示试验因素各水平组之间的差别大小是否有统计学意义。等效性试验主要用来显示试验因素各水平组之间的差别大小在临床上有无重要意义。优效性试验主要用来显示试验因素某水平组在观察指标上的取值情况是否优于其他组很多。非劣效性试验要用来显示试验因素某水平组在观察指标上的取值情况是否不比其他组差很多。当然，等效性检验、优效性检验、非劣效性检验在实际使用时都需要设定相应的界值。

(5) 把握变量性质。结果变量的性质对分析方法的选择有重大影响，设计类型相同的定量资料和定性资料，在选择分析方法时差别很大。以配对设计为例，若是一元定量资料，可选择配对设计一元定量资料的 t 检验(或符号秩和检验)；若是定性资料，则需采用一致性检验(2×2 列联表资料也可视分析目的选择 McNemar χ^2 检验)。此外，结果变量的个数对分析方法的选择也有很大影响。若同时考察的结果变量只有一个，则资料为某设计类型的一元定量(或定性)资料，若同时考察的结果变量有两个或两个以上，则资料为某设计类型的多元定量(或定性)资料，两者可选的分析方法有很大差别。

(6) 检查前提条件。很多统计分析方法在实际应用时，都有一定的使用条件。例如，成组设计一元定量资料，若使用成组设计定量资料 t 检验来分析，则需要资料满足独立性、正态性和方差齐性。其中，独立性是指一批实验数据彼此之间是相互独立的；正态性是指观测指标的取值服从或近似服从正态分布；方差齐性是指特定因素各水平组合条件下观测指标的取值所对应的总体方差相等或近似相等。同样，对于双向无序的 $R \times C$ 列联表资料而言，若采用 χ^2 检验来分析，则要求理论频数小于 5 的单元格数目未超过总格子数的 1/5(若为四格表资料，还同时要求总样本例数不小于 40)。

14.2 合理选择统计分析方法处理一元定量资料的关键技术

下面本节以一系列实例，来介绍如何正确选择统计分析方法。

【例 14-1】 人们普遍认为欧洲各国人均年消费酒量相当于纯酒精 8 L，为验证这种认识是否符合实际情况，研究者挑选了 10 个欧洲城镇，将每人每年平均消费的酒量换算成纯酒精数(升)，数据已经按升序排列：4.12, 5.81, 7.63, 9.74, 10.39, 11.92, 12.32, 12.89, 13.54, 14.45。

【分析】 研究者欲比较欧洲各国人均年消费酒量是否为 8 L 纯酒精，选择了 10 个欧洲城镇，计算得到这些城镇每人每年平均消费的纯酒精量，此资料为定量资料。从欧洲城镇中抽取了 10 个样本，以这 10 个城镇每人年平均消费的纯酒精量与 8 L 进行比较，属于单组设计一元定量资料。分析时，比较欧洲各国人均年消费酒量与 8 L 这一参考值之间有无统计学差异，属于差异性分析范畴。若资料服从正态分布，可采用单组设计定量资料的 t 检验分析此资料；反之，可采用符号秩和检验。

【例 14-2】 某医生欲观察胺碘酮治疗急性心肌梗死(AMI)并心房颤动(房颤)的临床疗效及安全性，选择 34 例 AMI 并房颤患者作为受试对象，首剂静脉注射胺碘酮 150mg，随后以 0.5 ~ 1.5mg/min 维持静脉滴注，静脉注射用药 3 ~ 7d。研究者观测了受试对象在治疗前和治疗后的心率情况，见表 14-1。请比较：用药前后，患者心率之间的差异是否有统计学意义？

【分析】研究者欲比较胺碘酮用药前后,患者心率之间的差异是否有统计学意义,从比较类型上来说,属于差异性试验范畴。选择 34 例 AMI 并房颤患者作为受试对象,记录他们在治疗前和治疗后的心率数据。由于每名患者都先后被观察了两次心率的情况,而心率为计数变量,所以该资料为配对设计一元定量资料。若患者治疗前后心率的差值进行正态性检验,若服从正态分布,可采用配对设计定量资料的 t 检验分析此资料;反之,可采用符号秩和检验。

【例 14-3】某研究者欲研究稳心颗粒对冠心病患者心率变异性的影响,选取 26 例冠心病(CHD)患者随机分为 2 组:稳心颗粒组(14 例)和对照组(12 例)。两组均接受常规剂量的阿司匹林(100mg, qd)及单硝酸异山梨酯(20mg, bid)作为基础治疗,稳心颗粒组加服稳心颗粒(9g, tid),治疗过程持续两周,观察并记录相关实验数据,其中两组受试对象个体 24h 连续正常 R-R 间期的标准差(SDNN)如下。试比较稳心颗粒对冠心病患者 SDNN 是否存在影响?

稳心颗粒组: 154.13 122.35 139.05 126.73 157.36 122.87 137.16 136.47 130.39 156.43 132.71 161.57 137.68 141.26

对照组: 100.34 88.57 97.79 98.67 114.37 130.48 124.95 84.72 96.81 101.62 124.32 92.15

【分析】研究者以基础治疗为对照,研究基础治疗加服稳心颗粒对冠心病患者心率变异性的影响,属于差异性试验范畴。观察指标为 SDNN,属于定量指标,因此该资料为成组设计一元定量资料。若资料服从正态分布并满足方差齐性要求,可采用成组设计定量资料的 t 检验分析此资料;若资料服从正态分布但不满足方差齐性要求,可采用成组设计定量资料的 t' 检验或 Wilcoxon 秩和检验分析此资料;若资料不服从正态分布,可采用 Wilcoxon 秩和检验。

【例 14-4】欲研究两种药物的生物等效性,研究者选取两组同质的受试者参与实验,每组 12 人,分别给予试验药和对照药,计算两组患者服药后 AUC(血药浓度曲线下面积)的值,结果如下。

试验组: 103.4 69.48 96.74 59.92 72.17 94.26 68.17 74.37 48.52 94.54 84.44 95.68

对照组: 90.11 58.21 89.3 77.71 101.3 97.22 77.71 79.84 61.62 97.51 96.06 85.8

【分析】研究者欲比较两种药物的生物等效性,属于等效性试验范畴。观察指标为 AUC,为定量指标,因此该资料为成组设计一元定量资料。分析时可采用成组设计一元定量资料的等效性检验进行处理,若资料不服从正态分布,可对其进行变量变换。美国食品与药品监督管理局建议,生物等效性评价可采用平均值比值的 80% ~ 120% 作为是否等效的判断标准,若对数据进行对数变换,对数转换后的 AUC 采用两平均值比值的 80% ~ 125% 作为是否等效的判断标准更优。

【例 14-5】某研究者欲探讨糖尿病、糖代谢异常患者颈动脉粥样硬化的情况,随机选取 34 例某院住院的患者,采血测定血糖及血总胆固醇等。根据血糖检查结果和美国糖尿病协会的分类标

表 14-1 16 名急性心肌梗死并心房颤动患者胺碘酮治疗前后心率值

患者编号	心率(次/分)		
	治疗前	治疗后	治疗前后差值
1	66	61	5
2	67	60	7
3	67	62	5
4	70	64	6
5	70	65	5
6	76	67	10
7	77	66	11
8	79	69	10
9	79	72	7
10	82	74	8
11	82	78	4
12	83	76	7
13	93	82	11
14	94	83	10
15	95	83	12
16	98	96	2

准将患者分为2型糖尿病组(10例)、糖代谢异常组(12例)和正常对照组(12例)。具体标准为,2型糖尿病组:空腹血糖 $>7\text{mmol/L}$ 或餐后2h血糖 $>11.1\text{mmol/L}$;糖代谢异常组:空腹血糖为 $6.1\sim 6.9\text{mmol/L}$;正常对照组:空腹血糖为 $3.9\sim 6.0\text{mmol/L}$,而餐后2h血糖 $<7.8\text{mmol/L}$ 。三组受试对象的血清总胆固醇结果如下,试比较他们之间血清总胆固醇含量的差异是否有统计学意义?

2型糖尿病组:1.64 1.71 2.13 2.28 1.58 1.78 3.24 0.75 3.25 2.05

糖代谢异常组:1.12 1.85 1.16 0.67 2.12 2.10 1.05 1.84 1.97 0.84 0.93 1.72

正常对照组:1.55 0.80 1.36 1.06 1.24 0.86 1.19 1.24 0.80 0.72 2.00 0.80

【分析】 研究者欲比较2型糖尿病患者、糖代谢异常患者和正常人的血清总胆固醇含量之间是否有差异,属于差异性试验的范畴。试验因素为“血糖水平”或“糖尿病类型或程度”(正常对照组即为无糖尿病),观察指标为“血清总胆固醇含量”,所以资料属于单因素三水平设计一元定量资料。若资料服从正态分布并满足方差齐性要求,可采用单因素三水平设计一元定量资料方差分析处理此资料;若资料服从正态分布但不满足方差齐性要求,可采用单因素三水平设计一元定量资料的Welch方差分析或Kruskal-Wallis秩和检验处理此资料;若资料不服从正态分布,可采用Kruskal-Wallis秩和检验。

【例14-6】 某研究者欲比较3种抗癌药物对小白鼠肉瘤的抑瘤疗效,先将15只染有肉瘤的小白鼠按体重大小配成5个区组,使每个区组内的3只小白鼠体重最接近,然后随机决定每个区组中的3只小白鼠接受3种药物的治疗,以肉瘤的重量为指标,实验结果见表14-2。试比较不同抗癌药物对小白鼠肉瘤的抑瘤效果之间的差别是否有统计学意义?

【分析】 研究者欲比较3种抗癌药物对小白鼠肉瘤的抑瘤疗效,属于差异性试验的范畴。试验因素为“抗癌药物种类”,观察指标为“肉瘤重量”,研究者认为体重是重要的非试验因素,故将其作为一个区组因素来安排试验,那么此资料就相当于在单因素三水平设计的基础上多考虑了一个区组因素的影响,对应的设计类型应为随机区组设计。在资料满足参数检验的前提条件时,可采用随机区组设计定量资料的方差分析,若资料不满足参数检验的前提条件,可采用Friedman秩和检验。

表14-2 不同药物作用后小白鼠肉瘤重量

区组	肉瘤重量(g)		
	*: A	B	C
1	0.82	0.65	0.51
2	0.73	0.54	0.23
3	0.43	0.34	0.28
4	0.41	0.21	0.31
5	0.68	0.43	0.24

注: * 代表“药物种类”

【例14-7】 某研究者欲探讨I型多聚ADP核糖合成酶(PARP1)在心肌梗死后大鼠心肌组织中的表达及活性变化,选择雄性Wistar大鼠36只,210~260g,随机等分为3组,分别接受腹腔注射PARP抑制剂3-氨基苯甲酰胺(3-AB)30mg/kg(低剂量3-AB干预组)、100mg/kg(高剂量3-AB干预组)或等体积的0.9%氯化钠溶液(AMI组)。3组大鼠均结扎冠状动脉左前降支(LAD)近端建立急性心肌梗死(AMI)模型,结扎术前15~20min及术后2h腹腔注射相应药物,以后每天注射同等剂量,直至处死当天为止。在术后1、3、7天时,均从每组中随机选取4只大鼠处死,处死前2h将大鼠以1.5%盐酸戊巴比妥钠10~20ml/kg腹腔内注射麻醉后仰卧固定,用超声仪检测大鼠左室收缩末期内径(LVESD)大小,结果见表14-3。

【分析】 研究者欲探讨3-氨基苯甲酰胺(3-AB)注射剂量和处死时间对大鼠左室收缩末期内径大小的影响,属于差异性试验的范畴。试验因素为“3-AB注射量”和“处死时间”,观察指标为“左室收缩末期内径大小”。表面上看,此设计像具有一个重复测量的两因素设计,这样看是把“处死时间”当成了重复测量因素。其实,不同时间点上获得的数据之间没有关联,不是来自于同一只大鼠,因为每只大鼠身上只能测到一个数据。所以,此资料不是具有一个重复测量

的两因素设计定量资料。由于试验条件为两因素各水平的全面组合，研究者在每种试验条件下均进行了 4 次独立重复试验，在没有专业依据认为这两个试验因素对观测结果的影响重要程度不同的情况下，此资料应为两因素析因设计定量资料。故应采用两因素析因设计定量资料的方差分析处理此资料。

表 14-3 各组大鼠术后第 1、3、7 天左室收缩末期内径大小

3-AB 注射 量 (mg/kg)	左室收缩末期内径 (cm)					
	时间 (d) : 1		3		7	
0	3.71	4.82	4.33	3.56	4.31	3.80
	4.02	3.84	4.14	3.58	4.18	4.11
30	2.58	2.18	2.69	2.69	3.92	3.64
	2.29	2.55	2.74	2.29	3.31	3.93
100	2.39	2.15	3.55	2.70	3.56	3.16
	2.63	2.82	2.46	3.69	3.36	3.13

【例 14-8】 某研究者欲观察心室再同步 (CRT) 起搏治疗慢性心力衰竭 (CHF) 的长期疗效，选择 15 例 CHF 患者，均有严重器质性心脏病，心功能 (NYHA 分级) Ⅲ ~ Ⅳ 级，所有患者心电图为窦性心律，ORS 波 > 130ms，左室舒张末期内径 (LVEDd) > 60mm，左室射血分数 (LVEF) < 35% 伴二尖瓣反流。15 例患者均植入 CRT 治疗，分别记录患者在植入前及植入后 1 个月、3 个月、6 个月、12 个月时左室舒张末期内径的大小，见表 14-4。

表 14-4 15 例 CHF 患者手术前后左室舒张末期内径变化

患者 编号	左室舒张末期内径 (mm)				
	观测时间：治疗前	植入 1 个月	植入 3 个月	植入 6 个月	植入 12 个月
1	67.8	64.4	59.8	59.4	58.8
2	59.0	56.8	55.1	50.6	42.8
3	89.8	85.9	81.4	81.3	77.4
4	60.7	62.0	59.2	58.6	53.8
5	68.3	65.5	62.6	64.4	57.4
6	80.6	78.5	78.1	73.8	71.2
7	75.8	75.4	68.1	66.8	67.1
8	90.5	89.8	87.8	85.0	83.8
9	77.7	74.2	73.8	69.0	68.7
10	78.0	77.1	73.7	71.6	68.3
11	80.2	78.6	80.1	76.9	70.0
12	63.8	64.7	58.2	58.6	57.4
13	83.1	83.3	80.8	80.9	74.6
14	72.0	70.5	65.3	65.3	65.5
15	89.6	87.7	82.5	82.0	79.2

【分析】 研究者欲探讨 CHF 患者在植入前及植入后 1 个月、3 个月、6 个月、12 个月时左室舒张末期内径的大小，属于差异性试验的范畴。对每一个患者来说，在 5 个时间点上分别被测量其左室舒张末期内径的大小，说明“观测时间”因素是一个重复测量的因素，且所有患者均接受同一种

治疗方法——植入 CRT，没有根据其他的试验因素对患者进行再分组，因而这是具有一个重复测量的单因素设计定量资料，应选用与其设计类型相对应的方差分析来处理数据。

【例 14-9】 某研究者欲观察隔日禁食结合耐饥食饵对小鼠生理指标的影响，将 20 只成年 ICR 小鼠随机等分为对照组（自由摄食标准饲料）、食饵组（自由摄食耐饥食饵）、隔日禁食组（隔日禁食标准饲料）、辟谷食饵组（隔日禁食耐饥食饵），试验 8 周，自由饮水，观察小鼠体重的变化，见表 14-5。请进行适当的统计分析。

表 14-5 4 组小鼠体重变化

摄食方式	饲料种类	编号	体重(g)			
			时间(周): 1	2	4	8
自由摄食	标准饲料	1	30.82	41.18	41.83	40.46
		2	29.03	35.11	38.86	38.84
		3	31.22	39.40	38.67	38.62
		4	32.35	42.08	45.66	42.35
		5	31.09	41.14	44.33	41.12
	耐饥食饵	1	32.38	40.32	41.45	39.61
		2	31.16	39.72	41.34	37.11
		3	28.99	36.61	36.94	34.27
		4	31.14	38.81	39.51	36.87
		5	32.04	40.89	43.52	40.04
隔日禁食	标准饲料	1	29.52	34.15	34.69	37.43
		2	31.62	35.61	36.55	40.56
		3	27.05	31.02	32.49	36.08
		4	31.22	34.61	35.24	37.72
		5	32.89	35.96	36.65	41.08
	耐饥食饵	1	33.55	35.04	35.89	38.88
		2	31.64	33.01	34.17	37.92
		3	28.04	30.42	32.82	34.46
		4	32.07	35.18	37.29	39.62
		5	31.59	32.21	33.40	36.70

【分析】 研究者欲隔日禁食结合耐饥食饵对小鼠体重的影响，属于差异性试验的范畴。对每一只小鼠来说，在 4 个时间点上分别被测量其体重，说明“时间”因素是一个重复测量的因素。此外，还有两个试验因素“摄食方式”（自由摄食或隔日禁食）和“饲料种类”（标准饲料或耐饥食饵），因而这是具有一个重复测量的三因素设计定量资料，应选用具有一个重复测量的三因素设计定量资料的方差分析来处理数据。

【例 14-10】 某研究者欲研究急性心肌缺血对犬左室心肌楔形组织块瞬时外向钾电流(I_{to})、跨壁复极离散度(TDR)变化的影响及其与室性心律失常的关系。选择 11 只成年杂种家犬，雌雄不限，体重 16~18kg，制备冠状小动脉灌流犬左室心肌楔形组织块电生理模型，应用浮置玻璃微电极和心电图同步记录技术，观察急性无灌流心肌缺血对内、中、外 3 层心肌细胞 I_{to} 、动作电位时程(APD)的影响。其中，APD 的变化以复极时的 APD(APD₉₀)计算，分别在模型建立成功后 0、5、10、20min 记录 APD 的值，实验结果见表 14-6。试进行合适的统计分析。

表 14-6 急性心肌缺血各阶段 3 层心肌细胞 APD 的变化

编号	APD ₉₀ (ms)											
	*:0 (外膜	中膜	内膜)	5 (外膜	中膜	内膜)	10 (外膜	中膜	内膜)	20 (外膜	中膜	内膜)
1	230.7	284.6	261.2	189.2	252.7	229.7	180.2	233.6	215.1	174.1	205.7	194.1
2	208.5	253.9	217.1	164.8	224.5	211.7	158.4	207.9	193.5	146.5	173.5	173.6
3	232.0	277.3	256.5	188.1	249.0	225.9	180.6	230.2	215.3	168.5	198.2	189.9
4	210.5	263.2	233.1	165.0	234.6	218.8	161.4	209.4	205.5	160.2	191.8	181.3
5	244.2	287.4	269.3	204.5	262.0	240.8	190.8	246.1	227.6	183.1	219.9	209.3
6	230.5	284.1	260.4	192.8	255.0	234.6	183.9	241.2	222.1	173.3	210.6	206.8
7	183.3	243.2	213.8	156.9	209.1	201.8	148.4	192.7	172.9	136.0	173.4	173.8
8	214.1	271.9	237.7	169.9	238.6	224.0	161.2	214.0	209.8	161.7	190.9	180.5
9	213.3	271.1	238.1	183.9	242.2	222.3	163.4	217.8	209.0	164.5	192.1	182.7
10	207.4	255.4	224.8	166.0	231.6	211.7	159.0	208.9	193.8	152.7	179.0	181.6
11	218.1	273.5	243.3	182.4	241.9	224.0	167.1	219.3	215.9	165.8	195.6	185.6

注：* 代表“时间×细胞层”

【分析】 研究者欲观察急性无灌流心肌缺血后不同时间上犬内、中、外 3 层心肌细胞 APD₉₀ 的变化情况，属于差异性试验的范畴。本研究涉及两个试验因素，即“观测时间”和“心肌细胞层种类”，前者有 4 个水平(0、5、10、20min)，后者有 3 个水平(外膜、中膜、内膜)。由于是对同一只家犬，在 4 个不同时间点上获得 3 层心肌细胞 APD₉₀ 的 12 个观测值，所以“观测时间”为重复测量因素。而“心肌细胞层种类”指同一只家犬左心室上具有可比性的 3 层心肌细胞，因而也是重复测量因素，所以此资料的试验设计类型为具有两个重复测量的两因素设计，应选用具有两个重复测量的两因素设计一元定量资料的方差分析来处理此资料。

14.3 合理选择统计分析方法处理一元定性资料的关键技术

【例 14-11】 为了调查某工厂产品生产合格率的情况，共对该工厂随机抽取 1000 件产品进行检验，发现其中 978 件合格，22 件不合格。已知业内相同产品的生产合格率为 98%，问该工厂产品合格率与业内平均水平有无差异？

【分析】 研究者欲比较某工厂产品生产合格率与同行业相同产品的合格率有无差异，属于差异性比较的范畴。研究者从工厂随机抽取 1000 件产品作为样本，所有观测对象均在一个组中，且给定了业内生产合格率的参考值(98%)，因此它属于单组设计一元定性资料。由于观测指标是二值变量(合格、不合格)，分析时可采用二项检验对数据进行统计处理。

【例 14-12】 某研究者欲研究高温作业人群高血压的患病情况，选择某钢厂高温作业工人 959 人，以同企业 257 名非高温作业工人为对照，分别测量两组人的血压情况。测量结果见表 14-7，欲分析两类作业工人的高血压患病率有无差异，应如何进行分析？

【分析】 研究者欲分析高温作业和非高温作业两类作业工人的高血压患病率有无差异，属于差异性比较的范畴。由于在调查中同时观察了受试对象的工作环境和是否患高血压，这一过程是在一很短的时间内调查获得的，且两个定性变量都只有 2 个水平，因此该资料为横断面研究设计的 2×2 表资料。若资料是满足

表 14-7 两组工人高血压患病情况

工作环境	例数(人)		
	*:是	否	合计
高温作业	171	788	959
非高温作业	27	230	257
合计	198	1018	1216

注：* 代表“是否患高血压”

一般 χ^2 检验的前提条件,则选用 χ^2 检验,否则,应根据资料情况选用校正的 χ^2 检验或 Fisher 精确检验。

【例 14-13】 某研究者欲研究中医药防治大肠癌术后复发转移的作用,选取 162 例大肠癌患者,分成两组:治疗组在西医放化疗的基础上给予中药治疗,服用相应中药汤药 3 个月,对照组仅给予西医放化疗。随访 3 年后,获得两组患者的复发转移情况,结果见表 14-8。请问:应如何分析两组患者复发转移率之间的差异是否有统计学意义?

【分析】 研究者欲分析两种方法(放化疗 + 中药、放化疗)治疗大肠癌后的复发转移情况有无统计学差异,属于差异性比较的范畴。研究者对受试对象施加处理后,随访 3 年,观察各组患者大肠癌的复发转移情况,施加处理在前,观察结果在后,且原因变量及结果变量均为 2 水平,所以此资料为队列研究设计的 2×2 表资料。欲比较两组患者复发转移率之间的差异是否有统计学意义,首先应将此资料视为横断面研究设计的 2×2 表资料进行处理。若得到两组患者复发转移率之间的差异有统计学意义时,再计算相对危险度 RR 值,然后采用 $MH\chi^2$ 检验来分析总体 RR 值与 1 之间的差异是否有统计学意义。

【例 14-14】 某研究组欲对人工流产与乳腺癌之间的关系进行探索性研究,选择某市 55 岁以下原发 I 期乳腺癌患者 1495 例,选择年龄频数与乳腺癌患者组相匹配的全人群妇女 1573 例。调查她们是否有人工流产史,结果见表 14-9。请问:应如何分析两组人群人工流产史之间的差异是否有统计学意义?

表 14-9 是否患乳腺癌与有无人工流产史情况

有无人工 流产史	例数(人)		
	是	否	合计
有	1003	1075	2078
无	492	498	990
合计	1495	1573	3068

注: * 代表“是否患乳腺癌”

两组人群人工流产史之间的差异是否有统计学意义,首先应将此资料视为横断面研究设计的 2×2 表资料进行处理。若得到两组人群人工流产史之间的差异有统计学意义时,再计算优势比 OR 值,然后采用 $MH\chi^2$ 检验来分析总体 OR 值与 1 之间的差异是否有统计学意义。

【例 14-15】 为观察复方薤白胶囊治疗慢性支气管炎急性发作期痰热郁肺证的疗效。随机选取 240 例符合要求的病人,均分为 2 组,试验组口服复方薤白胶囊,对照组口服芩暴红止咳胶囊,疗程均为 10d。两组疗效见表 14-10,如果试验组的有效率超过对照组 5%,才能认为复方薤白胶囊优于芩暴红止咳胶囊。试评价复方薤白胶囊治疗慢性支气管炎急性发作期痰热郁肺证的疗效是否优于芩暴红止咳胶囊。

【分析】 研究者欲评价复方薤白胶囊治疗慢性支气管炎急性发作期痰热郁肺证的疗效是否优于芩暴红止咳胶囊,属于优效性比较的范畴。此资料施加处理(口服复方薤白胶囊或芩暴红止咳胶

表 14-8 两种治疗方法随访 3 年
大肠癌的复发转移情况

治 疗 方 法	例数(人)		
	是	否	合计
放化疗 + 中药	17	67	84
放化疗	38	40	78
合计	55	107	162

注: * 代表“是否复发转移”

【分析】 研究者欲分析两组人群(原发 I 期乳腺癌妇女与非乳腺癌妇女)人工流产史之间的差异是否有统计学意义,属于差异性比较的范畴。由于是否患乳腺癌这一结果已经出现,据此向前追溯受试对象此前有无人工流产史,为由果溯因,且原因变量及结果变量均为 2 水平,所以此资料为病例-对照研究设计的 2×2 表资料。欲分析两

表 14-10 两组患者的治疗效果

药物种类	例数(人)		
	疗效: 有效	无效	合计
复方薤白胶囊	116	4	120
芩暴红止咳胶囊	111	9	120
合计	227	13	240

囊)在前,观察结果在后,为队列研究设计的 2×2 表资料。分析时,应采用成组设计定性资料的优效性检验。

【例 14-16】 某研究调查汉族和回族 ABO 血型分布情况,分别随机抽取宁夏某大学汉族和回族学生各 500 人,所得结果见表 14-11,试比较汉族和回族在血型分布上的差别有无统计学意义?

【分析】 研究者欲比较汉族和回族学生在血型分布上的差别有无统计学意义,属于差异性比较的范畴。由于原因变量为 2 值变量,结果变量均为多值名义水平,所以此资料为结果变量,为多值名义变量的 $2 \times C$ 表资料。若资料满足一般 χ^2 检验的前提条件,则选用 χ^2 检验,否则,应根据资料情况选用校正的 χ^2 检验或 Fisher 精确检验。

表 14-11 汉族和回族学生 ABO 血型分布情况比较

民 族	人 数				
	ABO 血型				
	A	B	O	AB	合计
汉族	135	146	159	60	500
回族	147	155	129	69	500
合计	282	301	288	129	1000

表 14-12 两种药物治疗心绞痛的疗效

治 疗 药 物	例数(人)			
	疗效:优	显效	无效	合计
益气通脉口服液	58	43	9	110
舒心口服液	32	61	16	109
合计	90	104	25	219

【例 14-17】 某研究组欲研究益气通脉

口服液治疗冠心病心绞痛的疗效,选择冠心病心绞痛气虚血瘀证性患者 110 例为试验组,予以益气通脉口服液治疗;另选病情相似患者 109 例为对照组,予以舒心口服液治疗。治疗 4 周后,观察两组患者疗效,结果见表 14-12,其中疗效优与显效合计为有效。若

欲了解两组患者治疗有效率之间的差异是否有统计学意义,应如何进行分析?若欲了解两组患者整体疗效之间的差异是否有统计学意义,应如何进行分析?

【分析】 研究者欲了解两组患者整体疗效之间的差异是否有统计学意义,属于差异性比较的范畴。本资料中原因变量为二值变量,结果变量为多值有序变量,所以此资料为结果变量,为多值有序变量的 $2 \times C$ 列联表资料。欲比较两组患者整体疗效之间的差异是否有统计学意义,可采用秩和检验或 Ridit 分析。若欲了解两组患者疗效分布之间的差异是否有统计学意义,可采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前提条件时使用)。

【例 14-18】 某研究者欲研究脑钠肽(BNP)检测对脓毒症患者血瘀证的临床诊断价值,选择已确诊的脓毒症患者 174 例,以 1988 年北京“血瘀证国际会议”制定的血瘀证诊断标准为金标准,将受试对象分为血瘀证组和非血瘀证组。另外,以 $\text{BNP} = 100\text{ng/L}$ 为截断点,小于此值为非血瘀证组,大于此值为

表 14-13 脓毒症 BNP 与血瘀证金标准诊断结果比较

BNP 法	例数(人)		
	金标准:血瘀证	非血瘀证	合计
血瘀证	86	35	121
非血瘀证	9	44	53
合计	95	79	174

血瘀证组。两种诊断方法诊断结果见表 14-13,若欲了解两种方法诊断结果不一致部分的差别是否有统计学意义,应如何进行分析?若欲了解两种方法的诊断结果是否一致,应如何进行分析?

【分析】 研究者欲了解两种方法诊断结果不一致部分的差别是否有统计学意义,以及两种方法的诊断结果是否一致,均属于差异性比较的范畴。试验中,研究者采用 BNP 法和金标准法同时对 174 例脓毒症患者进行诊断,判定脓毒症血瘀证和脓毒症非血瘀证。由于存在金标准,所以本资料应为特设金标准的配对设计四格表资料。欲了解两种方法诊断结果不一致部分的差别是否有统计学意义,可采用配对设计定性资料的 χ^2 检验,即 McNemar χ^2 检验;欲了解两种方法的诊断结果是否一致,则应采用一致性检验,即 Kappa 检验。

表 14-14 吸烟年限和龋齿的关系研究结果

吸 烟 年 限	例数		
	患	未患	合计
≤1	3	24	27
1~5	9	28	37
5~10	12	23	35
≥10	7	11	18
合计	31	86	117

注：* 代表“是否患龋属”

提条件时使用)比较此原因变量不同水平组的频数分布之间的差异是否有统计学意义外,还可采用线性趋势检验来分析原因变量与结果变量之间是否呈线性趋势。当然,若原因变量为多值名义变量,则此类资料只能采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前提条件时使用)来分析原因变量各水平组中结果变量构成之间的差异是否有统计学意义。

【例 14-20】 某大学对计算机专业、金融专业、传媒专业各 50 名学生进行心理学测试,并判断每个学生属于哪一种典型气质类型,所得结果见表 14-15。欲比较各专业学生的气质类型分布是否有统计学差异,应如何分析?

表 14-15 不同专业学生的四种气质类型分布

专 业	人 数				
	气质类型:多血质	胆汁质	抑郁质	黏液质	合计
计算机	16	13	7	14	50
金融	12	15	10	13	50
传媒	18	9	8	15	50
合计	46	37	25	42	150

【分析】 研究者欲比较各专业学生的气质类型分布是否有统计学差异,属于差异性比较的范畴。本资料中原因变量为多值名义变量,结果变量也是多值名义变量,所以本资料为双向无序的 $R \times C$ 列联表资料。欲比较各专业学生的气质类型分布是否有统计学差异,可采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前提条件时使用)。

【例 14-21】 某研究者欲观察健脾益肾方对小鼠移植性前胃癌术后肿瘤局部复发和肺转移的治疗作用,选取小鼠 56 只,制作前胃癌术后肺转移模型,然后随机分为三组。模型组 18 只,健脾益肾方治疗组 19 只,5-FU 化疗组 19 只。各组小鼠在原发瘤切除后,连续给药 15d,术后 15d 全部处死小鼠,检查肺转移情况,试验结果见表 14-16。欲分析各组肺转移程度之间的差异是否有统计学意义,应如何分析?

表 14-16 术后 15d 前胃癌小鼠肺转移情况

组 别	例数(只)				
	肺转移程度:0	I	II	III	合计
模型组	1	3	5	9	18
5-FU 化疗组	5	5	5	4	19

【例 14-19】 为了研究吸烟年限和龋齿的关系,某人搜集了表 14-14 所示的资料,试分析吸烟年限与龋齿患病率是否呈线性变化趋势。

【分析】 研究者欲分析吸烟年限和龋齿的关系,属于差异性比较的范畴。本资料中原因变量为多值名义变量,结果变量为二值变量,所以本资料为 $R \times 2$ 列联表资料。由于本资料中原因变量为多值有序变量,除可采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前

续表

组 别	例数(只)				
	肺转移程度: 0	I	II	III	合计
健脾益肾方治疗组	6	5	6	2	19
合计	12	13	16	15	56

【分析】 研究者欲分析各组前胃癌术后肺转移小鼠在肺转移程度上的差异是否有统计学意义,属于差异性比较的范畴。本资料中原因变量为多值名义变量,结果变量为多值有序变量,因此该资料为结果变量为有序变量的 $R \times C$ 列联表资料。分析各组肺转移程度之间的差异是否有统计学意义,可选用秩和检验或 Ridit 分析。

【例 14-22】 某研究者欲研究严重跟骨骨折手术时机的选择,选择跟骨骨折患者 71 例,分别于伤后 8h 内、8h~7d、7~14d 三个不同时期进行手术并比较手术疗效,结果见表 14-17。请问:应如何分析此资料?

表 14-17 三组患者术后疗效比较

手术时间	例数(人)				
	疗效: 优	良	可	差	合计
伤后 8h 内	17	12	2	0	31
伤后 8h~7d	5	6	5	1	17
伤后 7~14d	12	14	2	0	28
合计	34	32	9	1	76

【分析】 该资料中原因变量为“伤后手术时间”,结果变量为“疗效”,二者均为有序变量但本质属性不同,所以此资料为双向有序且属性不同的 $R \times C$ 列联表资料。分析此资料需结合列联表类型和分析目的来选择统计分析方法。若希望考察各组的频数分布是否相同,应将此资料视为双向无序的 $R \times C$ 列联表资料,采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前提条件时使用)。若希望考察各组间疗效的差异是否有统计学意义,应将此资料视为结果变量为有序变量的单向有序 $R \times C$ 列联表资料,选用秩和检验或 Ridit 分析。若希望考察两个有序变量之间的相关关系有无统计学意义,可采用 Spearman 秩相关分析。若希望进一步考察两变量之间的变化关系是否为直线关系时,可选用线性趋势检验。

【例 14-23】 某研究者欲评价腔内超声对直肠癌浸润深度诊断的准确性,选取 117 例未放化的直肠癌患者术前行腔内超声检查,参考 TNM 分期标准进行浸润程度分期诊断,并与术后病理组织学结果对照,试验结果见表 14-18。请问:应如何分析腔内超声检查与病理诊断结果是否一致?

表 14-18 两种诊断方法对 117 例直肠癌患者浸润深度的诊断结果

腔内超声 诊断分级	例数(人)				
	* : T ₁	T ₂	T ₃	T ₄	合计
T ₁	7	1	0	0	8
T ₂	1	15	10	0	26
T ₃	0	13	60	2	75

续表

腔内超声 诊断	例数(人)				
	* :T ₁	T ₂	T ₃	T ₄	合计
T ₄	0	0	0	8	8
合计	8	29	70	10	117

注：* 代表“病理诊断分级”

【分析】 研究者欲分析腔内超声检查与病理诊断结果是否一致，属于差异性比较的范畴。本资料中，原因变量和结果变量均为“直肠癌浸润深度”，是一有序变量，两变量性质相同且取值水平数及含义均相同，所以此资料应为双向有序且属性相同的列联表资料。研究者以术后病理组织学诊断为金标准，比较腔内超声检查与病理诊断结果是否一致，可采用一致性检验或 Kappa 检验。

【例 14-24】 某课题组欲评价如意金黄散外涂治疗不同程度的化疗药物性静脉炎的临床疗效，为短期内获得足够样本，分别由甲、乙两个医院抽取白血病住院病人化疗静脉给药而导致静脉炎患者，计 224 例。其中甲医院 120 例，男 68 例，女 52 例，年龄 18~72 岁，平均 51.6 岁，乙医院 104 例，男 64 例，女 40 例，年龄 23~69 岁，平均 50.9 岁，两医院患者一般情况比较，差异无显著性($P>0.05$)。根据 WHO 化疗毒性分级标准，0 级：无痛；Ⅰ级：无痛，但局部发红；Ⅱ级：轻度疼痛，局部发红；Ⅲ级：中度疼痛，局部轻度肿胀，灼热；Ⅳ级：重度顽固性疼痛，中、重度肿胀。结果见表 14-19。请问：若比较如意金黄散外涂治疗不同程度的化疗药物性静脉炎的临床疗效，应如何分析？

【分析】 此课题组欲评价如意金黄散外涂治疗不同程度的化疗药物性静脉炎的临床疗效之间有无统计学差异，属于差异性比较的范畴。试验中，研究者在甲、乙两个医院抽取不同程度的化疗药物性静脉炎患者作为受试对象，给予金黄散外涂治疗。结果变量为“疗效”，是一个二值变量，原因变量包括“不同医院”和“不同程度静脉炎”，前者为二值变量，后者为多值有序变量。所以，本资料应为结果变量为二值变量的高维列联表资料，可采用 CMH χ^2 检验、多重 Logistic 回归分析或对数线性模型进行处理。

表 14-19 如意金黄散外涂治疗不同程度的化疗药物性静脉炎及其疗效

不同医院	不同程度 静脉炎	例数(人)	
		疗效：有效	无效
甲医院	Ⅰ级	17	0
	Ⅱ级	40	0
	Ⅲ级	46	2
	Ⅳ级	12	3
乙医院	Ⅰ级	36	10
	Ⅱ级	18	10
	Ⅲ级	24	12
	Ⅳ级	8	12

【例 14-25】 某医生欲研究曼格磁贴治疗乳腺增生症的疗效，选取医院乳腺专科近两年内门诊就医的女性乳腺增生症患者 210 例，均经针吸细胞学活检排除良、恶性乳腺肿瘤。根据乳腺增生症诊断标准中的临床病理学分型将乳腺增生症分为乳腺组织增生(乳痛症)、乳腺腺病(小叶增长期、纤维腺病期、纤维化期)、乳腺囊肿病三类。210 例患者被分为两组：实验组 150 例，使用曼格Ⅱ号磁贴，该磁贴内容物为 SM 超强磁粉和括篓、红花、沉香、郁金、柴胡、丁香等中药。对照组 60 例，不使用磁粉，只使用中药，制成与曼格磁贴同样大小厚薄的软塑料片。3~5 d 换一次，3 个月为一个疗程，两组均治疗二个疗程。结果见表 14-20，请问：应如何分析曼格Ⅱ号磁贴疗效与不使用磁粉的同种贴剂疗效之间的差别是否具有统计学意义？

【分析】 此医生欲比较中药贴剂是否加磁在治疗乳腺增生症临床疗效上的差异是否有统计学意义，属于差异性比较的范畴。本资料结果变量为“疗效”，是一个有序变量，原因变量包括“增生症类型”和“磁粉用否”，前者为多值名义变量，后者为二值变量。所以，本资料应为结果变量为多值有序变量的高维列联表资料，可采用 CMH 校正的秩和检验或累积的多重 Logistic 回归分析进行处理。

表 14-20 中药贴剂是否加磁治疗乳腺增生症的治疗效果

增生症类型	磁 粉 用 否	例数(人)			
		疗效: 痊愈	显效	有效	无效
乳腺组织增生	否	0	2	4	12
	用	9	30	7	0
乳腺腺病	否	0	1	3	24
	用	4	48	15	5
乳腺囊肿病	否	0	0	4	10
	用	2	13	8	9

【例 14-26】 某研究者欲研究 Megsin 基因 E1-5‘UTR 区 A267G 与免疫球蛋白 A 型肾病(IgAN)阴虚证的相关性,筛选符合标准的原发性 IgAN 患者 120 例,进行 megsin 基因 E1-5‘UTR 区 A267G 的单核苷酸多态性(SNP)测序,观察中医证型与 SNP 的相关性。试验结果见表 14-21,请问:应如何分析不同性别不同中医证型 IgAN 患者中 SNP 分布的频数差异是否有统计学意义?

表 14-21 不同性别不同中医证型患者间 Megsin 基因 E1-5‘UTR 区 3 个 SNP 的分布差异

性 别	中 医 证 型	例数(人)		
		SNP: GG	GA	AA
男	气阴两虚	28	4	0
	肝肾阴虚	10	11	1
女	气阴两虚	35	5	0
	肝肾阴虚	10	14	2

【分析】 研究者欲分析不同性别不同中医证型 IgAN 患者中 SNP 分布的频数差异是否有统计学意义,属于差异性比较的范畴。本资料结果变量为“SNP”,是一个多值名义变量,原因变量包括“性别”和“中医证型”,均为二值变量。所以,本资料应为结果变量为多值名义变量的高维列联表资料,可采用扩展的多重 Logistic 回归分析或对数线性模型进行处理。

14.4 合理选择统计分析方法处理多元定量资料的关键技术

【例 14-27】 50 个白人男性申请警察局职位的体检数据(Gunst 和 Mason(1980)),指标包括:1. 身高(x_1 , cm); 2. 体重(x_2 , kg); 3. 肩宽 x_3 , cm); 4. 骨盆宽(x_4 , cm); 5. 最小胸围(x_5 , cm); 6. 大腿皮褶厚度(x_6 , mm); 7. 静息脉率(x_7); 8. 引体向上次数(x_8); 9. 最大肺活量(x_9 , kg); 10. 踏板跑步休息 5 分钟后脉率(x_{10}); 11. 最大踏板速度(x_{11}); 12. 体脂(x_{12})(资料来源: Dallas E. Johnson. Applied Multivariate Methods for Data Analysts)。其数据结构见表 14-22。若希望根据这些信息对 50 人进行分类,应如何分析?

表 14-22 50 个白人男性申请警察局职位的体检数据

id	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	179.6	74.20	41.7	27.3	82.4	19.0	64	2	158	108	5.5	11.91
2	175.6	62.04	37.5	29.1	84.1	5.5	88	20	166	108	5.5	3.13
3	166.2	72.96	39.4	26.8	88.1	22.0	100	7	167	116	5.5	16.89

续表

id	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
4	173.8	85.92	41.2	27.6	97.6	19.5	64	4	220	120	5.5	19.59
5	184.8	65.88	39.8	26.1	88.2	14.5	80	9	210	120	5.5	7.74
6	189.1	102.26	43.3	30.1	101.2	22.0	60	4	188	91	6.0	30.42
7	191.5	84.04	42.8	28.4	91.0	18.0	64	1	272	110	6.0	13.70
8	180.2	68.34	41.6	27.3	90.4	5.5	74	14	193	117	5.5	3.04
9	183.8	95.14	42.3	30.1	100.2	13.5	80	4	199	105	5.5	20.26
10	163.1	54.28	37.2	24.2	80.5	7.0	84	13	157	113	6.0	3.04
11	169.6	75.92	39.4	27.2	92.0	16.5	65	6	180	110	5.0	12.83
12	171.6	71.70	39.1	27.0	86.2	25.5	68	0	193	105	5.5	15.95
13	180.0	80.68	40.8	28.3	87.4	17.5	73	4	218	109	5.0	11.86
14	174.6	70.40	39.8	25.9	83.9	16.5	104	6	190	129	5.0	9.93
15	181.8	91.40	40.6	29.5	95.1	32.0	92	1	206	139	5.0	32.63
16	167.4	65.74	39.7	26.4	86.0	13.0	80	6	181	120	5.5	6.64
17	173.0	79.28	41.2	26.9	96.1	11.5	72	6	184	111	5.5	11.57
18	179.8	92.06	40.0	29.8	100.9	15.0	60	0	205	92	5.0	24.21
19	176.8	87.96	41.2	28.4	100.8	20.5	76	1	228	147	4.0	22.39
20	179.3	77.66	41.4	31.6	90.1	9.5	58	15	198	98	5.5	6.29
21	193.5	98.44	41.6	29.2	95.7	21.0	54	0	254	110	5.5	23.63
22	178.8	65.42	39.3	27.1	83.0	16.5	88	7	206	121	5.5	10.53
23	179.6	97.04	43.8	30.1	100.8	22.0	100	3	194	124	5.0	20.62
24	172.6	81.72	40.9	27.3	91.5	22.0	74	4	201	113	5.5	18.39
25	171.5	69.60	40.4	27.8	87.7	15.5	70	10	175	110	5.5	11.14
26	168.9	63.66	39.8	26.7	83.9	6.0	68	7	179	119	5.5	5.16
27	183.1	87.24	43.2	28.3	95.7	11.0	88	7	245	115	5.5	9.60
28	163.6	64.86	37.5	26.6	84.0	15.5	64	6	146	115	5.0	11.93
29	184.3	84.68	40.3	29.0	93.2	8.5	64	2	213	109	5.5	8.55
30	181.0	73.78	42.8	29.7	90.3	8.5	56	11	181	109	6.0	4.94
31	180.2	75.84	41.4	28.7	88.1	13.5	76	9	192	144	5.5	10.62
32	184.1	70.48	42.0	28.9	81.3	14.0	84	5	231	123	5.5	8.46
33	178.9	86.90	42.5	28.7	95.0	16.0	54	12	186	118	6.0	13.47
34	170.0	76.68	39.7	27.7	93.6	15.0	50	4	178	108	5.5	12.81
35	180.6	77.32	42.1	27.3	89.5	16.0	88	11	200	119	5.5	13.34
36	179.0	79.90	40.8	28.2	90.3	26.5	80	3	201	124	5.5	24.57
37	186.6	100.36	42.5	31.5	100.3	27.0	62	2	208	120	5.5	28.35
38	181.4	91.66	41.9	28.9	96.6	25.5	68	2	211	125	6.0	26.12
39	176.5	79.00	40.7	29.1	86.5	20.5	60	5	210	117	5.5	15.21
40	174.0	69.10	40.9	27.0	88.1	18.0	92	5	161	140	5.0	12.51
41	178.2	87.78	42.9	27.2	100.3	16.5	72	4	189	115	5.5	20.50
42	177.1	70.18	39.4	27.6	85.5	16.0	72	14	201	110	6.0	10.67
43	180.0	67.66	40.9	28.7	86.1	15.0	76	5	177	110	5.5	10.76
44	176.8	86.12	41.3	28.2	92.7	12.5	76	7	181	110	5.5	14.55
45	176.3	65.00	39.0	26.0	83.3	7.0	88	12	167	127	5.5	5.27
46	192.4	99.14	43.7	28.7	96.1	20.5	64	4	174	105	6.0	17.94
47	175.2	75.70	39.4	27.3	90.8	19.0	56	7	174	111	5.5	12.64
48	175.9	78.62	43.4	29.3	90.7	18.0	64	7	170	117	5.5	10.81
49	174.6	64.88	42.3	29.2	82.6	3.5	72	11	199	113	6.0	2.01
50	179.0	71.00	41.2	27.3	85.6	16.0	68	5	150	108	5.5	10.00

【分析】由表 24-22 的数据结构可知,全部申请者未按任何定性变量(即影响因素)进行分组,该资料应为单组设计多元定量资料。50 人为 50 个样品,根据他们在 11 个定量指标上的取值情况对其进行分类,可采用样品聚类分析。

【例 14-28】资料同【例 14-27】。若希望将多个性质不明的观测指标按其相似程度分为若干类,每一类反映事物的一种特性,应如何分析?

【分析】根据 50 个样品在 11 个观察指标上的取值情况,将这些观察指标按相似程度进行分类,可采用变量聚类分析。

【例 14-29】资料同【例 14-27】。若希望以互不相关的较少的综合指标(该综合指标是不能直接测量的)来反映原始指标所提供的大部分信息时,即希望找出“幕后”操纵原变量取值规律的隐变量时,应如何分析?

【分析】以少数综合指标反应原始指标所蕴含的主要信息,一般可采用主成分分析。

【例 14-30】资料同【例 14-27】。若希望用较少的互相独立的公共因子反映原有变量的绝大部分信息,即希望以最少的信息丢失将众多原有变量浓缩成少数几个具有一定命名解释性的公共因子,应如何分析?

【分析】主成分分析可得到少量的主成分,但这些主成分的含义常常较为难以解释。这时,可采用因子分析,对因子轴进行旋转,以得到可解释性较好的公因子。

【例 14-31】资料同【例 14-27】。研究者欲考察形态学指标(含身高、体重、肩宽、骨盆宽、最小胸围和大腿皮褶厚度)与功能指标(含静息脉率、最大肺活量、踏板跑步休息 5min 后脉率、最大踏板速度)之间的相关关系,试进行统计分析。

【分析】形态学指标和功能指标分别包含 6 个和 4 个具体的指标,也就是说这是两类或两组指标。研究它们之间的相关关系,应采用典型相关分析。它研究两组随机变量之间相关关系的一种多元统计分析方法,可将多个变量与多个变量之间的相关转化为少数几对新的综合变量之间的相关。

14.5 合理选择统计分析方法处理多元定性资料的关键技术

【例 14-32】某临床试验比较两种新器械与一种常规器械(A、B、C)对于关节痛的治疗效果。将 162 例关节痛患者分为 3 组, A 组 56 例使用常规器械; B 组 43 例使用试验器械 B; C 组 63 例使用试验器械 C。治疗效果见表 14-23。请分析器械与疗效之间的对应关系。

表 14-23 三组器械治疗效果比较

分 组	例 数				
	治疗效果:治愈	显著	改善	无效	合计
A	15	19	19	3	56
B	7	10	18	8	43
C	11	21	24	7	63
合计	33	50	61	18	162

【分析】将表 14-23 转化为数据库格式时,可发现此资料包含两列,一列为器械种类,一列为疗效,两列均为定性变量,所以此资料为二元定性资料。在研究器械与疗效之间的对应关系时,可采用定性资料的对应分析进行处理。

【例 14-33】某一项有关母婴的调查研究,对母亲的年龄、是否吸烟及其婴儿的情况进行了调查,其数据见表 14-24。请分析婴儿情况与母亲情况的对应关系。

表 14-24 母婴情况的调查数据

母 亲 年 龄	例 数(人)				
	是否吸烟	早产(1年内死亡)	1年后存活)	足月产(1年内死亡)	1年后存活)
年轻	否	50	315	24	4012
	是	9	40	6	459
年长	否	41	147	14	1594
	是	4	11	1	124

【分析】 将表 14-24 转化为数据库格式时，可发现此资料包含四列，分别为母亲年龄、母亲是否吸烟、婴儿是否足月出生、婴儿是否一年内存活，四列均为定性变量，所以此资料为四元定性资料。欲研究婴儿情况与母亲情况的对应关系时，可采用定性资料的对应分析进行处理。

参 考 文 献

[1] 高辉, 胡良平, 李长平等. 如何正确处理定性资料(一). 中西医结合学报, 2008, 6(10): 1086-1089.

[2] 高辉, 胡良平, 李长平等. 如何正确处理定性资料(二). 中西医结合学报, 2008, 6(11): 1194-1197.

[3] 胡良平. 科研设计与统计分析. 北京: 军事医学科学出版社, 2012: 279-732.

[4] 胡良平. SAS 统计分析教程. 北京: 电子工业出版社, 2010: 18-673.

[5] 胡良平. 心血管病科研设计与统计分析. 北京: 人民军医出版社, 2010: 50-172.

第 15 章 如何用 SAS 实现常用统计分析的关键技术

15.1 用 SAS 实现统计分析的简便快捷的方法

SAS 软件功能强大且很齐全，权威性高。使用方法有两种，可以采用菜单驱动法运行，但更灵活的使用方法是采用编程法运行。编程法效率高、灵活性强，但要求使用者熟练掌握 SAS 语言和一些高级编程技巧，这对初学者是一个很难逾越的障碍。

笔者在使用 SAS 软件多年之后，找到了一条有效的捷径：把有用的 SAS 引导程序汇集在一起，形成一个程序集，称为 SAS Pal 软件。用户只需根据此软件中的中文提示，阅读中文例题，用鼠标单击所对应的 SAS 程序文件名，就能将所需要的 SAS 引导程序调入 SAS 编程窗口。用户只需用自己的数据替换掉 SAS 引导程序的样例中的数据，将其发送给 SAS 系统执行，很快就能从输出窗口获得所需要的 SAS 输出结果。

下面以电子工业出版社《SAS 统计分析教程》这本书为例，简要介绍 SAS Pal 软件的使用方法，主界面见图 15-1。



图 15-1 《SAS 统计分析教程》SAS Pal 主界面

首先，用户需要根据界面中部上方的文字提示，下载并存储好数据文件。其次，在这个主界面的左下方，给出了与书本配套的各章节的目录。单击任一章节，右侧将出现该章节的例题及相应的 SAS 引导程序。以单击第 3 章为例，界面见图 15-2。

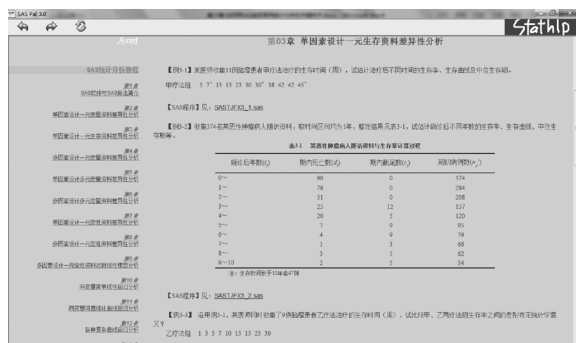


图 15-2 《SAS 统计分析教程》SAS Pal 第 3 章界面

最后，单击打开例题下方的 SAS 程序，即可在 SAS 软件中打开。用户根据需要将自己的数据替换掉程序中的数据，提交给 SAS 系统去执行，就能获得所需要的输出结果。这是初学者学会用 SAS 解决统计学问题最有效的方法。

15.2 用 SAS 实现一元定量资料的统计分析

【对例 14-1 的分析】 若资料服从正态分布，可采用单组设计定量资料的 t 检验分析此资料；反之，可采用符号秩和检验。SAS 程序如下，程序名为 jcyx15_1.sas。

<pre>data jcyx15_1; input x@@ ; cards; 4.12 5.81 7.63 9.74 10.39 11.92 12.32 12.89 13.54 14.45 ; run;</pre>	<pre>ods html; proc univariate mu0=8 normal; run; ods html close;</pre>
--	---

【对例 14-2 的分析】 若患者治疗前后心率的差值进行正态性检验，若服从正态分布，可采用配对设计定量资料的 t 检验分析此资料；反之，可采用符号秩和检验。SAS 程序如下，程序名为 jcyx15_2.sas。

<pre>data jcyx15_2; input x y@@ ; z=x-y; cards; 66 61 67 60 67 62 70 64 70 65 76 67 77 66 79 69 79 72 82 74 82 78 83 76 93 82 94 83 95 83 98 96 ; run;</pre>	<pre>ods html; proc univariate normal; var z; run; ods html close;</pre>
--	--

【对例 14-3 的分析】 若资料服从正态分布并满足方差齐性要求，可采用成组设计定量资料的 t 检验分析此资料；若资料服从正态分布但不满足方差齐性要求，可采用成组设计定量资料的 t' 检验或 Wilcoxon 秩和检验分析此资料；若资料不服从正态分布，可采用 Wilcoxon 秩和检验。SAS 程序如下，程序名为 jcyx15_3.sas。

<pre>data jcyx15_3; input g \$ n; do i=1 to n; input x@@ ; output; end; cards; wenxin 14 154.13 122.35 139.05 126.73 157.36 122.87 137.16 136.47 130.39 156.43 132.71 161.57 137.68 141.26 duizhao 12</pre>	<pre>ods html; proc univariate normal; var x; class g; run; proc ttest cochrans; var x; class g; run; proc nparlway wilcoxon; var x; class g;</pre>
---	---

```

100.34 88.57 97.79 98.67
114.37 130.48 124.95 84.72
96.81 101.62 124.32 92.15
;
run;

```

```

run;
ods html close;

```

【对例 14-4 的分析】 采用成组设计一元定量资料的等效性检验进行处理，若资料不服从正态分布，可对其进行变量变换。根据美国食品与药品监督管理局建议，对数据进行对数变换，对数转换后的 AUC 采用两平均值比值的 80% ~ 125% 作为是否等效的判断标准。SAS 程序如下，程序名为 jcyx15_4. sas。

```

data jcyx15_4;
do group='test','ref';
do i=1 to 12;
input y@@;
output;
end; end;
cards;
103.4 90.11 96.74 89.3 72.17
101.3 68.17 77.71 48.52 61.62
84.44 96.06 69.48 58.21 59.92
77.71 94.26 97.22 74.37 79.84
94.54 97.51 95.68 85.8
;
run;

```

```

ods html;
proc ttest data=jcyx15_4
dist=lognormal tost(0.8,1.25);
var y;
class group;
run;
ods html close;

```

【对例 14-5 的分析】 资料属于单因素三水平设计一元定量资料，若资料服从正态分布并满足方差齐性要求，可采用单因素三水平设计一元定量资料方差分析处理此资料；若资料服从正态分布但不满足方差齐性要求，可采用单因素三水平设计一元定量资料的 Welch 方差分析或 Kruskal-Wallis 秩和检验处理此资料；若资料不服从正态分布，可采用 Kruskal-Wallis 秩和检验。SAS 程序如下，程序名为 jcyx15_5. sas。

```

data jcyx15_5;
length group $ 10;
input group $ n;
do i=1 to n;
input x@@;
output;
end;
cards;
type2 10
1.64 1.71 2.13 2.28 1.58
1.78 3.24 0.75 3.25 2.05
yichang 12
1.12 1.85 1.16 0.67 2.12
2.10 1.05 1.84 1.97 0.84
0.93 1.72
zhengchang 12
1.55 0.80 1.36 1.06 1.24
0.86 1.19 1.24 0.80 0.72
2.00 0.80
;
run;

```

```

ods html;
proc univariate normal;
var x;
class group;
run;
proc glm;
class group;
model x=group/ss3;
means group/hovtest welch snk
dunnett('zhengchang');
run;
proc nparlway wilcoxon;
var x;
class group;
run;
ods html close;

```

【对例 14-6 的分析】 在资料满足参数检验的前提条件时,可采用随机区组设计定量资料的方差分析,若资料不满足参数检验的前提条件,可采用 Friedman 秩和检验。SAS 程序如下,程序名为 jcyx15_6. sas。

<pre>data jcyx15_6; do block=1 to 5; do drug=1 to 3; input x@@ ; output; end; end; cards; 0.82 0.65 0.51 0.73 0.54 0.23 0.43 0.34 0.28 0.41 0.21 0.31 0.68 0.43 0.24 ; run; proc univariate normal noprint data=jcyx15_6; var x; class block; output out=a normal=W probn=P; run; proc univariate normal noprint data=jcyx15_6; var x; class drug; output out=b normal=W probn=P; run;</pre>	<pre>ods html; data c; set a b; file print ods = (variables = (block drug W P)); title 'This is the results of Nor- mality test'; put _ods_; run; title; proc glm data=jcyx15_6; class block; model x=block/ss3; means block/hovtest; run; proc glm data=jcyx15_6; class drug; model x=drug/ss3; means drug/hovtest; run; proc glm data=jcyx15_6; class block drug; model x=block drug/ss3; means drug/snk; run; ods html close;</pre>
--	--

【对例 14-7 的分析】 此资料为两因素析因设计定量资料,应采用两因素析因设计定量资料的方差分析处理此资料。SAS 程序如下,程序名为 jcyx15_7. sas。

<pre>data jcyx15_7; do dose=1 to 3; do time=1 to 3; do mouse=1 to 4; input diameter@@ ; output; end; end; end; cards; 3.71 4.82 4.02 3.84 4.33 3.56 4.14 3.58 4.31 3.80 4.18 4.11 2.58 2.18 2.29 2.55 2.69 2.69 2.74 2.29 3.92 3.64 3.31 3.93 2.39 2.15 2.63 2.82 3.55 2.70 2.46 3.69 3.56 3.16 3.36 3.13 ; run;</pre>	<pre>ods html; proc glm; class dose time; model diameter=dose time/ss3; lsmeans dose* time/slice=dose; lsmeans dose* time/slice=time; lsmeans dose* time/tdiff pdiff; run; ods html close;</pre>
--	--

【对例 14-8 的分析】 这是具有一个重复测量的单因素设计定量资料,应选用与其设计类型相对应的方差分析来处理数据。SAS 程序如下,程序名为 jcyx15_8. sas。


```

data jcyx15_8;
  do patient=1 to 15;
    input m0 m1 m3 m6 m12@@ ;
    output;
  end;
  cards;
67.8 64.4 59.8 59.4 58.8
59.0 56.8 55.1 50.6 42.8
89.8 85.9 81.4 81.3 77.4
60.7 62.0 59.2 58.6 53.8
68.3 65.5 62.6 64.4 57.4
80.6 78.5 78.1 73.8 71.2
75.8 75.4 68.1 66.8 67.1
90.5 89.8 87.8 85.0 83.8
77.7 74.2 73.8 69.0 68.7
78.0 77.1 73.7 71.6 68.3
80.2 78.6 80.1 76.9 70.0
63.8 64.7 58.2 58.6 57.4
83.1 83.3 80.8 80.9 74.6
72.0 70.5 65.3 65.3 65.5
89.6 87.7 82.5 82.0 79.2
;
run;

ods html;
proc glm;
  model m0 m1 m3 m6 m12 = /nouni;
  repeated month 5 (0 1 3 6 12) /summary printe;
run;
ods html close;

```

【对例 14-9 的分析】 若患者治疗前后心率的差值进行正态性检验，若服从正态分布，可采用配对设计定量资料的 t 检验分析此资料；反之，可采用符号秩和检验。SAS 程序如下，程序名为 jcyx15_9.sas。

```

data jcyx15_9;
  do style=1 to 2;
    do forage=1 to 2;
      do subject=1 to 5;
        do time=1, 2, 4, 8;
          input y @@ ;
          output;
        end; end; end; end;
  cards;
30.82 41.18 41.83 40.46 29.03
35.11 38.86 38.84 31.22 39.40
38.67 38.62 32.35 42.08 45.66
42.35 31.09 41.14 44.33 41.12
32.38 40.32 41.45 39.61 31.16
39.72 41.34 37.11 28.99 36.61
36.94 34.27 31.14 38.81 39.51
36.87 32.04 40.89 43.52 40.04
29.52 34.15 34.69 37.43 31.62
35.61 36.55 40.56 27.05 31.02
32.49 36.08 31.22 34.61 35.24
37.72 32.89 35.96 36.65 41.08
33.55 35.04 35.89 38.88 31.64
33.01 34.17 37.92 28.04 30.42
32.82 34.46 32.07 35.18 37.29
39.62 31.59 32.21 33.40 36.70
;

proc mixed data=jcyx15_9;
  class style forage subject time;
  model y=style |forage |time;
  repeated/type = AR (1) sub = subject (style forage);
  ods output fitstatistics=d;
  ods output dimensions=d1;
run;
proc mixed data=jcyx15_9;
  class style forage subject time;
  model y=style |forage |time;
  repeated/type = SP (POW) (time)
  sub = subject (style forage);
  ods output fitstatistics=e;
  ods output dimensions=e1;
run;
% macro shuju (dataset, y);
data &dataset;
  set &dataset;
  rename value = &y;
run;
% mend shuju;
% shuju (a, VC) % shuju (b, CS)
% shuju (c, UN) % shuju (d, AR1)
% shuju (e, SP) % shuju (a1, VC)
% shuju (b1, CS) % shuju (c1, UN)

```

```

run;
proc mixed data=jcyx15_9;
  class style forage subject time;
  model y=style|forage|time;
  repeated/type = VC sub = subject
(style forage);
  ods output fitstatistics=a;
  ods output dimensions=a1;
run;
proc mixed data=jcyx15_9;
  class style forage subject time;
  model y=style|forage|time;
  repeated/type = CS sub = subject
(style forage);
  ods output fitstatistics=b;
  ods output dimensions=b1;
run;
proc mixed data=jcyx15_9;
  class style forage subject time;
  model y=style|forage|time;
  repeated/type = UN sub = subject
(style forage);
  ods output fitstatistics=c;
  ods output dimensions=c1;
run;

```

```

% shuju (d1, AR1) % shuju (e1, SP)
data f;
  merge a b c d e;
run;
data f1;
  merge a1 b1 c1 d1 e1;
run;
ods html;
proc print data=f;
  format _numeric_5.1;
run;
proc print data=f1;
run;
ods html close;

```

【对例 14-10 的分析】 此资料的试验设计类型为具有两个重复测量的两因素设计，应选用具有两个重复测量的两因素设计一元定量资料的方差分析来处理此资料。SAS 程序如下，程序名为 jcyx15_10. sas。

```

data jcyx15_10;
  do dog=1 to 11;
    do time=1 to 4;
      do layer=1 to 3;
        input y @@ ;
        output;
      end; end; end;
  cards;
230.7 284.6 261.2 189.2 252.7
229.7 180.2 233.6 215.1 174.1
205.7 194.1 208.5 253.9 217.1
164.8 224.5 211.7 158.4 207.9
193.5 146.5 173.5 173.6 232.0
277.3 256.5 188.1 249.0 225.9
180.6 230.2 215.3 168.5 198.2
189.9 210.5 263.2 233.1 165.0
234.6 218.8 161.4 209.4 205.5
160.2 191.8 181.3 244.2 287.4
269.3 204.5 262.0 240.8 190.8
246.1 227.6 183.1 219.9 209.3
230.5 284.1 260.4 192.8 255.0
234.6 183.9 241.2 222.1 173.3
210.6 206.8 183.3 243.2 213.8
156.9 209.1 201.8 148.4 192.7

```

```

proc mixed data=jcyx15_10;
  class dog time layer;
  model y=time|layer;
  repeated/type=VC sub=dog;
  ods output fitstatistics=a;
  ods output dimensions=a1;
run;
proc mixed data=jcyx15_10;
  class dog time layer;
  model y=time|layer;
  repeated/type=CS sub=dog;
  ods output fitstatistics=b;
  ods output dimensions=b1;
run;
proc mixed data=jcyx15_10;
  class dog time layer;
  model y=time|layer;
  repeated/type=AR(1) sub=dog;
  ods output fitstatistics=c;
  ods output dimensions=c1;
run;
ods html;
proc sql;
  select a.Descr, a.VALUE as VC,

```

```

172.9 136.0 173.4 173.8 214.1
271.9 237.7 169.9 238.6 224.0
161.2 214.0 209.8 161.7 190.9
180.5 213.3 271.1 238.1 183.9
242.2 222.3 163.4 217.8 209.0
164.5 192.1 182.7 207.4 255.4
224.8 166.0 231.6 211.7 159.0
208.9 193.8 152.7 179.0 181.6
218.1 273.5 243.3 182.4 241.9
224.0 167.1 219.3 215.9 165.8
195.6 185.6
;
run;

```

```

b.VALUE as CS, c.VALUE as AR1
from a, b, c
where a.Descr = b.Descr = c.Descr;
quit;
proc sql;
select a1.Descr, a1.VALUE as VC,
b1.VALUE as CS, c1.VALUE as AR1
from a1, b1, c1
where
a1.Descr = b1.Descr = c1.Descr;
quit;
ods html close;

```

15.3 用 SAS 实现一元定性资料的统计分析

【对例 14-11 的分析】 该资料为单组设计一元定性资料，分析时采用二项检验对数据进行统计处理。SAS 程序如下，程序名为 jcyx15_11.sas。

```

data jcyx15_11;
input group count;
cards;
1 978
2 22
;
run;

```

```

ods html;
proc freq;
tables group/binomial (p=0.98);
weight count;
run;
ods html close;

```

【对例 14-12 的分析】 该资料为横断面研究设计的 2×2 表资料。若资料是满足一般 χ^2 检验的前提条件，则选用 χ^2 检验，否则，应根据资料情况选用校正的 χ^2 检验或 Fisher 精确检验。SAS 程序如下，程序名为 jcyx15_12.sas。

```

data jcyx15_12;
do x=1 to 2;
do y=1 to 2;
input f@@;
output;
end; end;
cards;
171 788
27 230
;
run;

```

```

ods html;
proc freq;
tables x* y/chisq;
weight f;
run;
ods html close;

```

【对例 14-13 的分析】 此资料为队列研究设计的 2×2 表资料。分析时，首先应将此资料视为横断面研究设计的 2×2 表资料进行处理。若得到两组患者复发转移率之间的差异有统计学意义时，再计算相对危险度 RR 值，然后采用 $MH\chi^2$ 检验来分析总体 RR 值与 1 之间的差异是否有统计学意义。SAS 程序如下，程序名为 jeyx15_13.sas。

```

data jcyx15_13;
do x=1 to 2;
do y=1 to 2;

```

```

ods html;
proc freq;
tables x* y/chisq cmh;

```

```
input f@@ ;
output;
end; end;
cards;
17 67
38 40
;
run;

weight f;
run;
ods html close;
```

【对例 14-14 的分析】 此资料为病例-对照研究设计的 2×2 表资料。分析时，首先应将此资料视为横断面研究设计的 2×2 表资料进行处理。若得到两组人群人工流产史之间的差异有统计学意义时，再计算优势比 OR 值，然后采用 $MH\chi^2$ 检验来分析总体 OR 值与 1 之间的差异是否有统计学意义。SAS 程序如下，程序名为 jcyx15_14.sas。

```
data jcyx15_14;
do x=1 to 2;
do y=1 to 2;
input f@@ ;
output;
end; end;
cards;
1003 1075
492 498
;
run;

ods html;
proc freq;
tables x* y/chisq cmh;
weight f;
run;
ods html close;
```

【对例 14-15 的分析】 分析时，采用成组设计定性资料的优效性检验。SAS 程序如下，程序名为 jcyx15_15.sas。

```
data jcyx15_15;
do x=1 to 2;
do y=1 to 2;
input f@@ ;
output;
end; end;
cards;
116 4
111 9
;
run;

ods html;
proc freq;
tables x* y/chisq riskdiff (sup
margin=0.05);
weight f;
run;
ods html close;
```

【对例 14-16 的分析】 此资料为结果变量为多值名义变量的 $2 \times C$ 表资料。若资料是满足一般 χ^2 检验的前提条件，则选用 χ^2 检验，否则，应根据资料情况选用校正的 χ^2 检验或 Fisher 精确检验。SAS 程序如下，程序名为 jcyx15_16.sas。

```
data jcyx15_16;
do x=1 to 2;
do y=1 to 4;
input f@@ ;
output;
end; end;
cards;
135 146 159 60
147 155 129 69
;
run;

ods html;
proc freq;
tables x* y/chisq fisher;
weight f;
run;
ods html close;
```

【对例 14-17 的分析】 此资料为结果变量为多值有序变量的 $2 \times C$ 列联表资料。欲比较两组患者整体疗效之间的差异是否有统计学意义,可采用秩和检验。SAS 程序如下,程序名为 jcyx15_17.sas。

```
data jcyx15_17;
  do x=1 to 2;
    do y=1 to 3;
      input f@@ ;
      output;
    end; end;
  cards;
  58 43 9
  32 61 16
  ;
run;
```

```
ods html;
proc npar1way wilcoxon;
  var y;
  class x;
  freq f;
run;
ods html close;
```

【对例 14-18 的分析】 本资料为特设金标准的配对设计四格表资料。欲了解两种方法诊断结果不一致部分的差别是否有统计学意义,可采用配对设计定性资料的 χ^2 检验,即 McNemar χ^2 检验;欲了解两种方法的诊断结果是否一致,则应采用一致性检验,即 Kappa 检验。SAS 程序如下,程序名为 jcyx15_18.sas。

```
data jcyx15_18;
  do x=1 to 2;
    do y=1 to 2;
      input f@@ ;
      output;
    end; end;
  cards;
  86 35
  9 44
  ;
run;
```

```
ods html;
proc freq;
  tables x* y;
  weight f;
  test agree;
run;
ods html close;
```

【对例 14-19 的分析】 本资料为 $R \times 2$ 列联表资料。欲分析吸烟年限与龋齿患病率是否呈线性变化趋势,可采用 Cochran-Armitage 线性趋势检验。SAS 程序如下,程序名为 jcyx15_19.sas。

```
data jcyx15_19;
  do x=1 to 4;
    do y=1 to 2;
      input f@@ ;
      output;
    end; end;
  cards;
  3 24
  9 28
  12 23
  7 11
  ;
run;
```

```
ods html;
proc freq;
  tables x* y/trend;
  weight f;
run;
ods html close;
```

【对例 14-20 的分析】 本资料为双向无序的 $R \times C$ 列联表资料。欲比较各专业学生的气质类型分布是否有统计学差异,可采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前提条件时使用)。SAS 程序如下,程序名为 jcyx15_20.sas。

```
data jcyx15_20;
  do x=1 to 3;
    do y=1 to 4;
      input f@@ ;
      output;
    end; end;
  cards;
  16 13 7 14
  12 15 10 13
  18 9 8 15
  ;
run;
```

```
ods html;
proc freq;
  tables x* y/chisq;
  weight f;
run;
ods html close;
```

【对例 14-21 的分析】 本资料为结果变量为有序变量的 $R \times C$ 列联表资料。分析各组肺转移程度之间的差异是否有统计学意义，可选用秩和检验。SAS 程序如下，程序名为 jcyx15_21.sas。

```
data jcyx15_21;
  do x=1 to 3;
    do y=1 to 4;
      input f@@ ;
      output;
    end; end;
  cards;
  1 3 5 9
  5 5 5 4
  6 5 6 2
  ;
run;
```

```
ods html;
proc nparlway wilcoxon;
  var y;
  class x;
  freq f;
run;
ods html close;
```

【对例 14-22 的分析】 本资料为双向有序且属性不同的 $R \times C$ 列联表资料，分析此资料需结合列联表类型和分析目的来选择统计分析方法。若希望考察各组的频数分布是否相同，应将此资料视为双向无序的 $R \times C$ 列联表资料，采用 χ^2 检验(满足 χ^2 检验应用前提条件时使用)或 Fisher 精确检验(不满足 χ^2 检验应用前提条件时使用)。若希望考察各组间疗效的差异是否有统计学意义，应将此资料视为结果变量为有序变量的单向有序 $R \times C$ 列联表资料，选用秩和检验或 Ridit 分析。若希望考察两个有序变量之间的相关关系有无统计学意义，可采用 Spearman 秩相关分析。若希望进一步考察两变量之间的变化关系是否为直线关系时，可选用线性趋势检验。SAS 程序如下，程序名为 jcyx15_22.sas。

```
data jcyx15_22;
  do x=1 to 3;
    do y=1 to 4;
      input f@@ ;
      output;
    end; end;
  cards;
  17 12 2 0
  5 6 5 1
  12 14 2 0
  ;
run;
ods html;
/* 考察各组间频数分布有无差异*/
```

```
/* 考察各组间整体疗效有无差异*/
proc nparlway wilcoxon;
  var y;
  class x;
  freq f;
run;
/* 考察两变量间有无相关性*/
proc corr spearman;
  var y;
  with x;
  freq f;
run;
ods html close;
```

```
proc freq;
  tables x* y/chisq;
  weight f;
  exact fisher/mc alpha=0.05;run;
```

【对例 14-23 的分析】 本资料为双向有序且属性相同的列联表资料。研究者以术后病理组织学诊断为金标准，比较腔内超声检查与病理诊断结果是否一致，可采用一致性检验或称 Kappa 检验。SAS 程序如下，程序名为 jcyx15_23. sas。

<pre>data jcyx15_23; do x=1 to 4; do y=1 to 4; input f@@ ; output; end; end; cards; 7 1 0 0 1 15 10 0 0 13 60 2 0 0 0 8 ; run;</pre>	<pre>ods html; proc freq; tables x* y; weight f; test agree; run; ods html close;</pre>
--	---

【对例 14-24 的分析】 本资料应为结果变量为二值变量的高维列联表资料，可采用多重 Logistic 回归分析进行处理。SAS 程序如下，程序名为 jcyx15_24. sas。

<pre>data jcyx15_24; do x=1 to 2; do y=1 to 4; do z=1 to 2; input f@@ ; output; end; end; end; cards; 17 0 40 0 46 2 12 3 36 10 18 10 24 12 8 12 ; run;</pre>	<pre>ods html; proc logistic; model z = x y; weight f; run; ods html close;</pre>
--	---

【对例 14-25 的分析】 本资料为结果变量为多值有序变量的高维列联表资料，可采用累积的多重 Logistic 回归分析进行处理。SAS 程序如下，程序名为 jcyx15_25. sas。

<pre>data jcyx15_25; do x=1 to 3; do y=1 to 2; do z=1 to 4; x1=0; x2=0; if x=1 then x1=1;</pre>	<pre>ods html; proc logistic; weight f; model z = x1 x2 y; run; ods html close;</pre>
---	---

```
else if x=2 then x2=1;
input f@@ ;
output;
end; end; end;
cards;
0 2 4 12
9 30 7 0
0 1 3 24
4 48 15 5
0 0 4 10
2 13 8 9
;
run;
```

【对例 14-26 的分析】 本资料为结果变量为多值名义变量的高维列联表资料，可采用扩展的多重 Logistic 回归分析进行处理。SAS 程序如下，程序名为 jcyx15_26. sas。

```
data jcyx15_26;
do x=1 to 2;
do y=1 to 2;
do z=1 to 3;
input f@@ ;
output;
end; end; end;
cards;
28 4 0
10 11 1
35 5 0
10 14 2
;
run;

ods html;
proc logistic;
weight f;
model z (ref = '1 ') = x y/link =
glogit;
run;
proc logistic;
weight f;
model z (ref = '1 ') = y/link =
glogit;
run;
ods html close;
```

15.4 用 SAS 实现多元定量资料的统计分析

【对例 14-27 的分析】 根据 50 个白人男性申请者在 11 个定量指标上的取值情况对他们进行分类，可采用样品聚类分析。SAS 程序如下，程序名为 jcyx15_27. sas。

```
data jcyx15_27;
input x1 -x12;
cards;
179.6 74.2 41.7 27.3 82.4
19 64 2 158 108 5.5 11.91
175.6 62.04 37.5 29.1 84.1
5.5 88 20 166 108 5.5 3.13
.....
179 71 41.2 27.3 85.6 16
68 5 150 108 5.5 10
;
run;

ods html;
proc cluster outtree=tree
method = cen std ccc pseudo sim-
ple;
var x1 -x12;
run;
proc tree horizontal;
run;
ods html close;
```

【对例 14-28 的分析】 根据 50 个白人男性申请者在 11 个观察指标上的取值情况，将这些观察指标按相似程度进行分类，可采用变量聚类分析。SAS 程序如下，程序名为 jcyx15_28. sas。


```

data jcyx15_28;
input x1 -x12;
cards;
  179.6  74.2  41.7  27.3  82.4
19  64  2  158  108  5.5  11.91
  175.6  62.04  37.5  29.1  84.1
5.5  88  20  166  108  5.5  3.13
.....
  179  71  41.2  27.3  85.6  16
68  5  150  108  5.5  10
;
run;

```

```

ods html;
proc varclus;
  var x1 -x12;
run;
ods html close;

```

【对例 14-29 的分析】 以少数综合指标反应原始 11 个观察指标所蕴含的主要信息，可采用主成分分析。SAS 程序如下，程序名为 jcyx15_29.sas。

```

data jcyx15_29;
input x1 -x12;
cards;
  179.6  74.2  41.7  27.3  82.4
19  64  2  158  108  5.5  11.91
  175.6  62.04  37.5  29.1  84.1
5.5  88  20  166  108  5.5  3.13
.....
  179  71  41.2  27.3  85.6  16
68  5  150  108  5.5  10
;
run;

```

```

ods html;
proc princomp;
  var x1 -x12;
run;
ods html close;

```

【对例 14-30 的分析】 以少数综合指标反应原始 11 个观察指标所蕴含的主要信息，也可采用因子分析，通过对因子轴进行旋转，得到可解释性较好的公因子。SAS 程序如下，程序名为 jcyx15_30.sas。

```

data jcyx15_30;
input x1 -x12;
cards;
  179.6  74.2  41.7  27.3  82.4
19  64  2  158  108  5.5  11.91
  175.6  62.04  37.5  29.1  84.1
5.5  88  20  166  108  5.5  3.13
.....
  179  71  41.2  27.3  85.6  16
68  5  150  108  5.5  10
;
run;

```

```

ods html;
proc factor method=ml nfact=4
  rotate=varimax reorder score
  out=factorscore;
run;
proc print;
  var factor1 - factor4;
run;
ods html close;

```

【对例 14-31 的分析】 形态学指标和功能指标分别包含 6 个和 4 个具体的指标，也就是说这是两类或两组指标。研究它们之间的相关关系，可采用典型相关分析。SAS 程序如下，程序名为 jcyx15_31.sas。

```

data jcyx15_31;
input x1 -x12;
cards;
  179.6  74.2  41.7  27.3  82.4
19  64  2  158  108  5.5  11.91

```

```

ods html;
proc cancorr red;
  var x1 -x6;
  with x7 x9 x10 x11;
run;

```

175.6 62.04 37.5 29.1 84.1 5.5 88 20 166 108 5.5 3.13 179 71 41.2 27.3 85.6 16 68 5 150 108 5.5 10 ; run;	ods html close;
---	-----------------

15.5 用 SAS 实现多元定性资料的统计分析

【对例 14-32 的分析】 研究器械与疗效之间的对应关系，可采用定性资料的对应分析进行处理。SAS 程序如下，程序名为 jcyx15_32. sas。

data jcyx15_32; do x='A','B','c'; do y='治愈','显著','改善','无 效'; input f@@; output; end; end; cards; 15 19 19 3 7 10 18 8 11 21 24 7 ; run;	ods html; ods graphics on; proc corresp freqout; weight f; tables x, y; run; ods graphics off; ods html close;
---	---

【对例 14-33 的分析】 研究婴儿情况与母亲情况的对应关系，可采用定性资料的对应分析进行处理。SAS 程序如下，程序名为 jcyx15_33. sas。

data jcyx15_33; do x='年轻','年长'; do y='不吸烟','吸烟'; do z='早产','足月'; do t='1 年内死亡','1 年内存活'; input f@@; output; end; end; end; end; cards; 50 315 24 4012 9 40 6 459 41 147 14 1594 4 11 1 124 ; run;	ods html; ods graphics on; proc corresp mca; tables x y z t; weight f; run; ods graphics off; ods html close;
---	--

参 考 文 献

[1] 高辉, 胡良平, 李长平等. 如何正确处理定性资料(一). 中西医结合学报, 2008, 6(10): 1086-1089.
[2] 高辉, 胡良平, 李长平等. 如何正确处理定性资料(二). 中西医结合学报, 2008, 6(11): 1194-1197.
[3] 胡良平. 科研设计与统计分析. 北京: 军事医学科学出版社, 2012: 279-356.
[4] 胡良平. SAS 统计分析教程. 北京: 电子工业出版社, 2010: 2-193.
[5] 胡良平. 心血管病科研设计与统计分析. 北京: 人民军医出版社, 2010: 50-144.

附录 A 胡良平统计学专著及配套软件简介

1. 《医学统计学与 SAS 应用技巧》简介

胡良平, 周士波主编. 北京: 中国科学技术出版社, 15.67 万字, 1991(定价: 4.2 元). 本书基于 DOS 版 SAS 6.03 软件, 介绍 SAS 应用入门、医学试验设计、常用统计分析、多元统计分析和 VAX SAS 应用入门。

2. 《医学统计应用错误的诊断与释疑》简介

胡良平主编. 北京: 军事医学科学出版社, 17.8 万字, 1999(定价 12 元). 本书针对医学科研和医学期刊中常犯的统计学错误, 讲解如何识别错误, 如何正确选用统计分析方法。

3. 《医学统计学内容概要、考题精选与考题详解》简介

胡良平编著. 北京: 军事医学科学出版社, 37 万字, 2000(定价 22 元). 本书简明扼要地概述了医学统计学的主要内容, 精选出 20 套适合检查统计学应用水平的考题, 并附有详细的解答。

4. 《现代统计学与 SAS 应用》简介

胡良平主编. 北京: 军事医学科学出版社, 1996, 2000, 2002(定价 40 元). 本书详细地介绍各种常用和多元统计分析方法, 并给出了手工计算和用 6.04 版 SAS 软件实现统计计算的方法和结果的解释。

5. 《Windows SAS 6.12 & 8.0 实用统计分析教程》简介

胡良平编著. 北京: 军事医学科学出版社, 96.9 万字, 2001(定价 52 元). 本书不仅介绍各种常用和多元统计分析方法, 还着重介绍 Windows SAS 6.12 & 8.0 的使用方法(含编程法和非编程法), 详细介绍辨析多因素设计类型的技巧和用 SAS 实现实验设计的方法。

6. 《医学统计学基础与典型错误辨析》简介

胡良平, 李子建主编. 北京: 军事医学科学出版社, 60.4 万字, 2003(定价 36 元). 本书详细地介绍学习统计学的策略、所必需的基本知识、常用的描述性统计分析方法和假设检验方法。

7. 《检验医学科研设计与统计分析》简介

胡良平主编. 北京: 人民军医出版社, 64 万字, 2004(定价 65 元). 本书紧紧围绕实验设计的三要素和四原则、分析定量资料和定性资料的要领、诊断性试验和一致性检验中的统计分析方法等重要内容, 从正反两方面详细阐述了学习和灵活运用这些知识的方法和技术。

8. 《医学统计实用手册》简介

胡良平主编. 北京: 人民卫生出版社, 48.5 万字, 2004(定价 30 元). 鉴于目前医学科研和医学期刊中存在大量误用和滥用统计学的现象, 本书通过分析这些现象产生的根源和实质, 有针对性地提出了解决这些问题的对策。

9. 《统计学三型理论在实验设计中的应用》简介

胡良平主编. 北京: 人民军医出版社, 50.1 万字, 2006(定价 45 元). 本书针对“许多人学了多

遍统计学仍不得要领,几乎是一用就错”的普遍现象,提出了彻底解决的对策,其精髓就是“统计学三型理论(简称‘三型理论’)”,即统计学问题基本上都可归结为“表现型”、“原型”和“标准型”,准确把握每个具体问题中的“三型”,将能科学合理地解决科研工作中与统计学有关的实际问题。事实上,统计学中的全部内容皆可运用“三型理论”来解说,但本书仅关注“科研设计”,特别是“实验设计”方面的问题。

10. 《医学统计实战练习》简介

胡良平主编.北京:军事医学科学出版社,83.4万字,2007(定价66元).本书收录笔者21年来从事统计教学、科研、咨询和培训工作中积累的各种考试真题以及根据审稿的稿件和公开发表的论文中提取的资料改编而成的新题,总共约有1000余道,并给出了每一道题的详细解答。

11. 《口腔医学科研设计与统计分析》简介

胡良平主编.北京:人民军医出版社,54万字,2007(定价65元).书中给出了取自口腔医学科研设计和统计分析的大量实例,运用“统计学三型理论”辨析“实验设计、统计描述和统计分析”中出现的错误,在给出正确做法的同时,还给出了带有原始数据的各种实例,用SAS软件演示统计分析的全过程和部分手工计算过程。还给出了估计样本含量的公式、实例和用SAS实现计算的方法。

12. 《统计学三型理论在统计表达与描述中的应用》简介

胡良平主编.北京:人民军医出版社,55.3万字,2008(定价80元).本书运用统计学三型理论,透过各种具体的统计表达和描述方面问题的“表现型”,揭示其“原型”,进而将“原型”正确地转变为“标准型”,使统计表达与描述方面的问题尽可能得到圆满解决。

13. 《科研课题的研究设计与统计分析(第一集)》简介

胡良平主编.北京:军事医学科学出版社,72.5万字,2008(定价55元).本书取材于我国2006年500多种生物医学期刊中影响因子较高的23种期刊,查阅这些期刊中近3000篇论著,从中挑选出具有广泛代表性的论著约300篇,主要从统计研究设计和统计分析方法选用两个方面,来剖析论著中存在的统计学问题,从而提示出我国生物医学科研工作质量需要进一步提高。

14. 《医学统计学——运用三型理论分析定量与定性资料》简介

胡良平主编.北京:人民军医出版社,72.3万字,2009(定价115元).本书在统计学思想指导下,运用统计学三型理论,透过各种具体科研问题所呈现的“表现型”,揭示其“原型”,进而将“原型”正确地转变为“标准型”,全面系统地介绍各种实验设计类型下收集的定量与定性资料的假设检验方法以及用SAS软件实现统计计算和结果解释。除常用的定量与定性资料的统计分析外,还介绍了META分析方法和高维列联表资料的各种处理方法。

15. 《科研课题的研究设计与统计分析(第二集)》简介

胡良平主编.北京:军事医学科学出版社,69.5万字,2009(定价56元).针对科研工作者所写出的学术论文和硕士与博士研究生所写出的学位论文在统计学方面存在很多问题的现实,本书全面介绍撰写高质量的论文所必需掌握的科研设计知识、统计分析知识和国际著名统计分析系统(SAS软件)使用知识,并针对生物医学科研领域中一些主干学科的特点,分析了约15个主干学科硕士和博士研究生学位论文中存在的统计学错误。从正反两个方面,揭示科研设计和统计分析的重要性,有利于提高科研工作者和研究生的科研素质、科研质量和论文的水平。

16. 《医学统计学——运用三型理论进行多元统计分析》简介

胡良平主编.北京:人民军医出版社,41.0万字,2010(定价70.0元).本书涵盖了现代多元统

计分析方法中的绝大部分内容,运用三型理论为指导,对多元统计分析方法进行了科学的分类,有利于实际工作者学习和使用。其内容包括变量聚类分析、主成分分析和探索性因子分析、典型相关分析、结构方程模型分析、无序样品聚类分析和有序样品聚类分析、多维尺度分析、各种设计定量资料的多元方差分析和多元协方差分析、判别分析、对应分析及其 SAS 实现。

17. 《心血管病科研设计与统计分析》简介

胡良平主编. 北京:人民军医出版社,47.5 万字,2010(定价 60 元). 本书内容分正反两个方面,正面讲述统计学中的主要内容,包括统计表达与描述、实验设计、定量与定性资料统计分析、简单相关回归分析和多重回归分析;围绕这些内容,又针对人们误用统计学的实际案例,对差错进行辨析与释疑。无论是正面还是反面内容,基本上都取材于与心血管疾病有关的我国数十种学术期刊中的科研论文。

18. 《SAS 统计分析教程》简介

胡良平主编. 北京:电子工业出版社,106.5 万字,2010(定价 68 元). 本书内容丰富且新颖,实用面宽且可操作性强。涉及定量与定性资料差异性和预测性分析、变量间和样品间相互与依赖关系及近似程度分析、数据挖掘与基因表达谱分析、绘制统计图与实验设计、SAS 语言和 SAS 非编程模块用法。这些内容高质量高效率地解决了实验设计、统计表达与描述、各种常用和多元统计分析、现代回归分析和数据挖掘、SAS 语言基础和 SAS 实现及结果解释等人们迫切需要解决却又十分棘手的问题。

19. 《SAS 实验设计与统计分析》简介

胡良平主编. 北京:人民卫生出版社,88.8 万字,2010(定价 72 元). 本书内容涉及面十分宽泛,由 SAS 软件基础、SAS 非编程模块介绍、SAS 编程法用法介绍、SAS 高级编程技术及其应用和 SAS 语言基础 5 篇组成,涵盖了 SAS 软件及其语言的基础和高级用法,实验设计、统计表达与描述和统计分析的主要内容及 SAS 实现。

20. 《医学统计学——运用三型理论进行现代回归分析》简介

胡良平主编. 北京:人民军医出版社,45.2 万字,2010(定价 75 元). 本书介绍现代回归分析方法中的大部分内容,包括多重线性回归分析、岭回归分析、各种复杂曲线回归分析、主成分回归分析、Poisson 回归分析、Probit 回归分析、负二项回归分析、配对和非配对设计定性资料多重 Logistic 回归分析、对数线性模型分析、生存分析和时间序列分析。

21. 《医学遗传统计分析与 SAS 应用》简介

胡良平,郭晋主编. 北京:人民卫生出版社,41.3 万字,2011(定价 36 元). 本书结合实例,介绍如何用 SAS 实现四大类遗传数据的统计分析方法,并介绍简明遗传学的概念与原理、遗传资料统计分析的原理。

22. 《正确实施科研设计与统计分析——统计学三型理论的应用与发展》简介

胡良平主编. 北京:人民军医出版社,87.8 万字,2011(定价:139 元). 本书全面介绍如何在三型理论指导下进行科研设计、统计表达与描述、常用统计分析、现代回归分析、多元统计分析和 SAS 实现方法。科研设计部分涵盖了概念、要点、设计类型等;统计表达与描述部分涵盖了统计表、统计图和概率分布等;常用统计分析部分涵盖了一元定量与定性资料的差异性分析;现代回归分析部分涵盖了包括多重线性回归分析、生存分析和时间序列分析等十余种现代回归分析方法;多元统计分析部分涵盖了包括变量聚类分析、判别分析和对应分析等十余种现代多元

统计分析方法;以上各部分均涉及如何用 SAS 软件巧妙实现的技术和方法,并有配套软件 SAS-Pal 方便程序调用。

23.《中医药科研设计与统计分析》简介

胡良平,王琪主编.北京:人民卫生出版社,41.4 万字,2011(定价 36.00 元).本书结合中医药领域中科研实例,不仅从正面介绍试验设计、统计表达与描述、统计分析方法及 SAS 实现技术,还对实际工作者在运用前述内容过程中所犯的各种错误进行了辨析与释疑。

24.《临床科研设计与统计分析》简介

胡良平,陶丽新主编.北京:中国中医药出版社,70.7 万字,2012(定价 45.00 元).本书主要对临床科研设计与统计分析问题进行阐述,同时还用较大篇幅揭示了临床科研课题和论文中的统计学错误,并给出了辨析与释疑。全书中的统计计算均用 SAS 软件实现。

25.《面向问题的统计学——(1)科研设计与统计分析》简介

胡良平主编.北京:人民卫生出版社,119.1 万字,2012(定价 98.00 元).本书分为 6 篇共 54 章,内容涉及消除学习统计学时的心理顾虑、统计思想、三型理论、科研设计、质量控制、表达与描述、单因素设计一元定量与定性资料统计分析、单组设计二元定量资料相关与回归分析和 SAS 语言基础与高级编程技术。

26.《面向问题的统计学——(2)多因素设计与线性模型分析》简介

胡良平主编.北京:人民卫生出版社,97.5 万字,2012(定价 80.00 元).本书分为 6 篇共 52 章,内容涉及多因素试验设计类型及其定量与定性资料的差异性分析和现代回归分析、判别分析、生存分析和时间序列分析;还介绍多水平模型分析法和综合分析法。

27.《面向问题的统计学——(3)试验设计与多元统计分析》简介

胡良平主编.北京:人民卫生出版社,85.2 万字,2012(定价 65.00 元).本书分为 5 篇共 25 章,内容涉及三类典型的多元数据结构(单组设计多元定量资料、单因素多水平设计多元定量资料、相似或不相似度矩阵)的各种多元统计分析方法,其代表性方法有主成分分析、样品聚类分析、对应分析、多维尺度分析、多元方差和协方差分析。

28.《外科科研设计与统计分析》简介

胡良平,毛玮主编.北京:中国协和医科大学出版社,40 万字,2012(38 元).本书分为三篇,第 1 篇统计学内容概要,包括统计表达与描述、试验设计、定量与定性资料的统计分析、简单相关与回归分析、多重线性回归分析与多重 Logistic 回归分析;第 2 篇外科科研中常见统计学错误辨析与释疑;第 3 篇医学统计学要览,以“问题引导”的形式提纲挈领地介绍“科研设计要览”与“统计分析要览”。

29.《科研设计与统计分析》简介

胡良平主编.北京:军事医学科学出版社,130.5 万字,2012(98 元).本书用 7 篇 31 章约 130 多万字的篇幅,概述了国内外迄今为止应该涵盖在统计学之中的绝大部分精彩内容:富含唯物辩证法精髓和心理学分析的统计思想、使统计思想具体化并具有可操作性的三型理论、灵活运用三型理论解决科研设计、统计表达与描述、各种简单与复杂统计分析、用国际著名统计分析系统 SAS 实现与前述全部内容有关的计算、结果解释和结论陈述。

30.《呼吸系统科研设计与统计分析》简介

胡良平,鲍晓蕾主编.北京:军事医学科学出版社,53.8 万字,2013.1(55 元).本书以近几年

出版的与呼吸科研相关杂志为主要的资料来源,在阐述统计学的基本理论、知识和技能的基础上,突出培养统计学思维方法、科研设计能力和应用统计分析方法的能力,以及计算机在处理临床科研资料中的正确应用技术。书中还用较大篇幅介绍呼吸科研课题和论文中常见统计学错误案例的辨析与释疑、SAS 软件的基础知识和使用技巧。

31.《护理科研设计与统计分析》简介

胡良平,关雪主编.北京:军事医学科学出版社,47.7 万字,2013.1(50 元).本书以近几年出版的护理科研相关杂志为主要的资料来源,在阐述统计学的基本理论、知识和技能的基础上,突出培养统计学思维方法、科研设计能力和应用统计分析方法的能力,以及计算机在处理护理科研资料中的正确应用技术。书中自始至终采用“识别错误”、“正确引导”和“归纳总结”的写作思路,把人们实施护理课题和撰写学术论文中常出现的错误呈现出来,并逐一对错差进行辨析与释疑;对案例所涉及的统计学基础知识进行系统梳理,从正面加以引导;对有关的统计理论和方法,从原理上进行归纳总结,以便实际工作者不仅知其然,还能知其所以然。

32.《脑血管病科研设计与统计分析》简介

胡良平,贾元杰主编.北京:军事医学科学出版社,50.3 万字,2013.5(58 元).本书结合脑血管病临床科研实际,比较全面地介绍从事临床科研工作所必需的思维方法、统计学基础理论和基本的统计分析技术,内容包括统计思想与三型理论在脑血管病科研中的应用、脑血管病科研基础——统计表达与描述、脑血管病科研设计、脑血管病试验设计、脑血管病临床试验设计、脑血管病调查设计、样本量估计与检验效能分析、常见多因素试验设计类型辨析、定量与定性资料统计分析、简单相关与回归分析、多重线性回归分析与多重 Logistic 回归分析。

33.《临床试验设计与统计分析》简介

胡良平,陶丽新主编.北京:军事医学科学出版社,54.7 万字,2013.5(58 元).本书结合临床科研和临床试验实际,首先介绍临床前研究和临床研究的主要内容,不仅在内容的安排上达到了承上启下的效果,使读者很自然地进入临床试验的情境之中;而且在要点把握上也起到了言简意赅、纲举目张的作用,使读者能在尽可能短的时间内领悟和抓住临床试验的核心和要领。在此基础上,结合笔者在国家级新药评审中发现的诸多问题,揭示了新药或医疗器械临床试验研究中的陷阱和识别错误的策略;介绍如何把握好临床试验研究中的三要素、四原则、设计类型和比较类型的概念、方法和技术要领。进而针对临床试验研究中使用频率最高的设计类型——成组设计,围绕四种比较类型、定量与定性资料、假设检验、样本量和检验效能估计等关键性问题,结合临床实例逐一进行介绍,并对同类问题进行了比较研究。

34.《非线性回归分析与 SAS 智能化实现》简介

胡良平,高辉主编.北京:电子工业出版社,51.5 万字,2013.7(定价 39 元).本书概述了回归分析的概念、分类、简单直线、曲线回归分析和多重线性回归分析、复杂固定模式和非固定模式曲线回归分析、单水平和多水平多重曲线回归分析。每种回归分析方法,都介绍分析目的、数据结构(问题与数据)、切入点(分析与解答)、统计模型(计算原理)、分析步骤(含 SAS 实现)。在固定模式单水平非线性回归分析中,涉及的统计模型有二项型和三项型指数曲线模型、Logistic 和 Gompertz 和 Richards 生长曲线模型、Bleasdale-Nelder 和 Halliday 和 Farazdaghi-Harris 产量-密度曲线模型;在非固定模式单和多水平多重非线性回归分析中,涉及的统计模型有二值结果变量定性资料单、多水平 Logistic、probit、互补双对数回归模型;多值有序结果变量定性资料单、多水平累积 Logistic、Probit、互补双对数回归模型;多值名义结果变量定性资料单和多水平扩展 Logistic 回归模型、计数

资料单和多水平 Poisson 和负二项回归模型。在上述各种情况下,还给出了同类问题的比较研究和 SAS 智能化实现及结果解释。

35.《课题设计与数据分析—关键技术与标准模板》简介

胡良平主编. 北京:军事医学科学出版社,48.5 万字,2014.1(48 元)。本书以“如何做好科研课题”为出发点和落脚点,开门见山,直奔主题,第 1 章从正反两种不同的视角,全面介绍了课题设计的基本概念、关键技术、具体做法和常见错误的辨析与释疑;第 2 章介绍了智源临床研究执行平台,它是一个智能化很高的数据管理和数据分析软件平台,它集数据网络平台录入、随机分组、逻辑核查、与国际著名统计分析软件 SAS 实现无缝对接等功能于一身;第 3 章和第 4 章介绍了临床试验研究中不可缺少的两个关键技术,即样本含量估计和随机化的 SAS 实现;第 5 章介绍了临床试验数据管理的标准操作规程、质量控制、具体流程和建立数据库的多款软件;第 6 章到第 9 章介绍了与高质量完成科研课题密切有关的第二部分内容,即对资料的统计表达描述和各种统计分析。与众不同的是,书中所介绍的统计分析方法几乎都可采用 SAS 智能化实现,免去了使用者在分析过程中很多不必要的担心和麻烦。

36.《SAS 语言基础与高级编程技术》简介

胡良平,胡纯严主编. 北京:电子工业出版社,73.3 万字,2014.5(定价 59 元)。本书全面系统地介绍了国际著名的统计分析系统 SAS 软件的主要内容,包括 SAS 语言基础、SAS 高级编程技术、SAS 9.2 和 9.3 版新增内容及用法简介、用 SAS 实现试验设计及处理病态数据的两个过程简介,其中前两部分是本书的重点。SAS 语言基础部分涵盖了如下内容:SAS 软件介绍、导入访问外部数据、基本 SAS 语言及其应用、常用 SAS 函数及其应用;而 SAS 高级编程技术部分包括如下内容:宏及其应用、SQL 及其应用、ODS 及其应用、数组 (ARRAY) 及其应用、IML 及其应用和如何掌握 SAS 语言的核心技术。书中还介绍了 SAS 9.2 和 9.3 版中一些新增过程和新增选项以及部分实用新过程的使用方法和技巧。

37.《医学综合统计设计与数据分析》简介

胡良平,赵铁牛,李长平主编. 北京:电子工业出版社,59.8 万字,2014.5(定价 45 元)。本书作者查阅了医学综合类期刊数十种,以每种期刊近 5 年中刊载的学术论文为研究对象,概括地总结了各类期刊论文中统计学应用的现况和在统计学应用方面存在的问题,先从正面讲解在生物医学和临床科研领域内如何正确使用统计学的理论、技术和方法,又列举了各种期刊论文中经常出现的误用和滥用的典型案例,对案例中出现的差错进行辨析与释疑,从正反两个方面启发和引导广大读者正确运用统计学、提高识别统计学错误的能力。内容涉及统计设计、统计表达与描述、统计分析、SAS 实现、结果解释和结论陈述。最后,又以一章的篇幅,提纲挈领、言简意赅地总结了学习统计学的方法。

38. 配套软件简介

(1)与前述介绍的《现代统计学与 SAS 应用》和《Windows SAS 6.12 & 8.0 实用统计分析教程》两本书对应的 SAS 引导程序,即 SAS Pal 软件由李子建研制,需要者可发电子邮件联系:lphu812@sina.com。

(2)与前述介绍的《口腔医学科研设计与统计分析》一书对应的 SAS 引导程序,即 SAS Pal 软件由胡纯严研制,需要者可发电子邮件联系:valenccia@sina.com 或 lphu812@sina.com。

(3)与《统计学三型理论在统计表达与描述中的应用》一书对应的 SAS 引导程序,即 SAS Pal 软件由胡纯严研制,需要者或有疑问者请发电子邮件联系:valenccia@sina.com。

(4) 与《统计学三型理论在定量与定性资料统计分析中的应用》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(5) 与《医学统计学——运用三型理论进行多元统计分析》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(6) 与《医学统计学——运用三型理论进行现代回归分析》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(7) 与《正确实施科研设计与统计分析——统计学三型理论的应用与发展》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(8) 与《SAS 统计分析教程》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(9) 与《科研设计与统计分析》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(10) 与《非线性回归分析与 SAS 智能化实现》一书对应的 SAS 引导程序, 即 SAS Pal 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

—— 统计分析教材 ——

SAS统计分析应用 (第2版) (董大钧)

SAS统计分析教程 (胡良平)

SAS语言基础与高级编程技术 (胡良平)

非线性回归分析与SAS智能化实现 (胡良平)

SPSS统计分析方法及应用 (第3版) (薛薇)

SPSS统计分析 (第4版) (卢纹岱)

SPSS 19 (中文版) 统计分析实用教程 (邓维斌)

应用统计与SPSS应用 (朱红兵)

PASW/SPSS Statistics中文版统计分析教程 (第3版)
(李志辉)

SPSS简明教程 (第10版) (何丽娟译)

社会调查方法与实践 (杜智敏)

社会科学统计方法 (第4版) (朱红兵译)

SPSS Modeler数据挖掘方法及应用 (第2版)
(薛薇)

MATLAB & Excel定量预测与决策

——运作案例精编 (张建林)

MATLAB定量决策五大类问题

——50个运作管理经典案例分析 (张建林)

医学综合统计设计与数据分析 (胡良平)

★基础医学统计设计与数据分析 (胡良平)

推荐出版及购书热线

Tel:010-88254531

E-mail:qinshl@phei.com.cn

Fax:010-88254560



欢迎登录 **免费** 获取本书教学资源
http://www.hxedu.com.cn

统计分析教材



ISBN 978-7-121-22793-6

董大钧



ISBN 978-7-121-10976-8

胡良平



ISBN 978-7-121-22989-3

胡良平



ISBN 978-7-121-20995-6

胡良平



ISBN 978-7-121-18949-4

薛薇



ISBN 978-7-121-10580-7

卢纹岱



ISBN 978-7-121-16132-2

邓维斌



ISBN 978-7-121-12760-1

朱红兵



ISBN 978-7-121-11944-6

李志辉



ISBN 978-7-121-21010-5

陈文沛



即将出版
杜智敏



ISBN 978-7-121-14219-2

朱红兵



ISBN 978-7-121-22203-0

薛薇



ISBN 978-7-121-17750-7

张健林



ISBN 978-7-121-19514-3

张健林



ISBN 978-7-121-23004-2

胡良平



ISBN 978-7-121-23194-0

胡良平



策划编辑：秦淑灵

责任编辑：郝黎明

责任美编：孙焱津

ISBN 978-7-121-23194-0



9 787121 231940 >

定价：45.00 元